



**Legal
Wings**
institute

Autor
Juliano Maranhão



Professor livre-docente da Faculdade de Direito da Universidade de São Paulo, sócio titular da Maranhão & Menezes e fundador, ao lado de professores da Poli/USP e IME/USP, da Lawgorithm Research Association. Fundador e Diretor do Legal Wings Institute.

Pesquisadora Assistente
Beatriz Sousa



Bacharel pela Faculdade de Direito da Universidade de São Paulo (FDUSP), com dupla graduação francesa pela Université de Lyon. Advogada associada no Maranhão & Menezes. Pesquisadora de direito e tecnologia na Associação Lawgorithm e no Instituto Legal Wings.

SUMÁRIO

1.	INTRODUÇÃO E ESPECIFICAÇÃO DO OBJETO DESTE ESTUDO	4
1.1.	Contexto.....	4
1.2.	Síntese do argumento.....	7
1.3.	Especificação do objeto deste estudo	9
2.	A OBRA AUTORAL E SUA PROTEÇÃO.....	15
2.1.	A obra	15
2.2.	A reprodução ou utilização da obra: analógica versus digital	18
2.3.	O treinamento de modelos de IAG não implica uso expressivo da obra individual.....	20
3.	INADEQUAÇÃO DO MODELO AUTORAL PARA REMUNERAÇÃO PELO USO DE OBRAS NO TREINAMENTO DE MODELOS DE IAG.....	29
3.1	Dimensão qualitativa dos custos do modelo proposto pelo PL 2338/23	29
3.1.1.	a obrigação de listar obras protegidas	29
3.1.2.	a obrigação de gerir o consentimento de obras identificadas	30
3.1.3.	a obrigação de remunerar autores ou detentores de direitos autorais.....	31
3.2	Implicações negativas para a cultura nacional e para o desenvolvimento de IAGs no Brasil	32
4.	CONCLUSÃO.....	35
5.	REFERÊNCIAS.....	37

1. INTRODUÇÃO E ESPECIFICAÇÃO DO OBJETO DESTE ESTUDO

1.1. Contexto

Sistemas de inteligência artificial generativa (IAG) são capazes de realizar uma grande variedade de tarefas que simulam o exercício de criatividade humana,¹ como a elaboração de textos, com notável fluência e coerência, a manipulação de atributos visuais para geração de imagens realistas e a composição de elementos de áudio para gerar músicas e simular vozes, permitindo uma série de explorações criativas por artistas, no campo da literatura, música, artes visuais e audiovisuais e por prestadores de serviços, que se utilizam desses sistemas para obter associações que permitem novas dimensões de análise e auxiliam o processo criativo.

Tais sistemas têm experimentado rápido crescimento em aplicações e em número de usuários, desde o surgimento de geradores de imagem, passando pela popularização de Assistentes de IA, que integram funcionalidades multimodais (de geração de texto, áudio, imagens e vídeo). Por reduzirem significativamente o custo de produção de conteúdo intelectual e permitirem novas explorações criativas, IAGs prometem revolucionar a economia, com ganhos de produtividade em diversos setores movidos por criação de conteúdo², como, por exemplo, as aplicações no setor jurídico para automação na elaboração de peças, contratos, *e-discovery* e *due diligences*,³ no campo do jornalismo para apoio nas atividades de redação, organização e análise de informações⁴; ou no setor de publicidade, com

¹ ALEMANHA. Federal Office for Information Security. *Generative AI Models: Opportunities and Risks for Industry and Authorities*. Disponível em: https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/KI/Generative_AI_Models.pdf?__blob=publicationFile&v=4> Acesso em 20 de jul. de 2024. COLUMBIA ENGINEERING, *Artificial Intelligence (AI) vs. Machine Learning*, Disponível em: <https://ai.engineering.columbia.edu/ai-vs-machine-learning/>. GOOGLE, *Artificial Intelligence (AI) vs. Machine Learning*, Disponível em: <https://cloud.google.com/learn/artificial-intelligence-vs-machine-learning>. IBM, *What is generative AI?*, Disponível em: <https://www.ibm.com/think/topics/generative-ai>. Acessos em 26 out. 2025

² MCAFEE, Andrew. *Generally Faster: The Economic Impact of Generative AI*. Abril 2024. Disponível em: https://storage.googleapis.com/gweb-uniblog-publish-prod/documents/Generally_Faster_-_The_Economic_Impact_of_Generative_AI.pdf. Acesso em 27 out. 2025. MCKINSEY & COMPANY. *The economic potential of generative AI: The next productivity frontier*. Junho 2023. Disponível em: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier?os=vpmkn75tqhpmkpsxtq&>. Acesso em 27 out. 2025.

³ BROWN, Dylan. *Tracking the irresistible rise of generative AI*. LexisNexis. Fevereiro 2024. Disponível em: <https://www.lexisnexis.co.uk/blog/future-of-law/tracking-the-irresistible-rise-of-generative-ai>. MORAN, Lyle. *54% of in-house legal professionals support generative AI usage*. Thomson Reuters Legal Dive. Junho 2023. Disponível em: <https://www.legaldive.com/news/thomson-reuters-generative-ai-legal-use-cases-in-house-legal-corporate-tax/654455/>. Acessos em 27 out. 2025.

⁴ Como notam Dodds et al, ferramentas de IA já estavam presentes no jornalismo antes do “boom” da IAG, como ferramentas de transcrição de voz em texto desde os anos 1990- mas o setor também experimentou expectativas de ganhos de produtividade e novas formas de automação a partir do lançamento do ChatGPT e a onda de ferramentas de IAG em 2022. DODDS, Tomás, YEUNG, Wang N., MELLADO, Claudia, LIMA-SANTOS, Mathias-Felipe. *On Controlled Change: Generative AI's Impact on*

o emprego das capacidades multimodais da IAG.⁵

Estima-se, atualmente, que sistemas de IAG, embora ainda em seus primeiros anos, atingiram um valor de mercado de cerca de 16 bilhões de dólares em 2024, com a perspectiva de alcançar 109 bilhões até 2030.⁶ Esse mercado tem recebido investimentos significativos (superando 56 bilhões de dólares em 2024)⁷ e tem gerado novas pressões competitivas nos chamados “mercados digitais”,⁸ em que diversos *players* buscam incorporar funcionalidades de IA em seus serviços *online* (e.g. aceleração dos investimentos em IA pelo Google e incorporação de sínteses nos resultados do buscador, ou o Adobe Photoshop, que passou a incorporar geradores de imagem).

O desenvolvimento de sistemas de IAG depende do treinamento de modelos com base em vastos conjuntos de dados extraídos de fontes diversas disponíveis na internet, frequentemente por meio de processo de mineração de dados (*text and data mining*)⁹, técnicas automatizadas de coleta (*“data scraping”*) ou pela aquisição de grandes bases de dados por agentes desenvolvedores. Essa prática tem suscitado questões quanto à proveniência dos dados, que envolve a legitimidade do acesso e a licitude de seu uso para fins de treinamento de modelos fundacionais de IA, uma vez que podem incluir obras protegidas por direitos autorais, conteúdos que exprimem aspectos da personalidade — como imagem, voz e estilo — ou ainda dados pessoais sensíveis.¹⁰

Professional Authority in Journalism. Outubro 2025. DOI:10.48550/arXiv.2510.19792. Acesso em 04 nov. 2025.

⁵ WORLD FEDERATION OF ADVERTISERS *apud* MOONSHOT NEWS. *Majority using GenAI for marketing despite concerns about technology*. Setembro 2024. Disponível em: <https://moonshot.news/news/advertising-news/majority-using-genai-for-marketing-despite-concerns-about-technology/>. INTERNATIONAL ADVERTISING BUREAU. *Generative AI Playbook for Advertising*. Disponível em: https://www.iab.com/wp-content/uploads/2025/01/IAB_GenerativeAIPlaybook_January_26.pdf. Acessos em 27 out. 2025.

⁶ GRAND VIEW RESEARCH. *Generative AI Market (2025 - 2030)*. Disponível em: <https://www.grandviewresearch.com/industry-analysis/generative-ai-market-report>. Acesso em 04 nov. 2025.

⁷ S&P GLOBAL. *GenAI funding hits record in 2024 boosted by infrastructure interest*. Janeiro 2025. Disponível em: <https://www.spglobal.com/market-intelligence/en/news-insights/articles/2025/1/genai-funding-hits-record-in-2024-boosted-by-infrastructure-interest-87132257>. Acesso em 04 nov. 2025.

⁸ MARANHÃO, Juliano et al. *Competition in AI Markets*. Legal Wings Institute, setembro 2025. Disponível em: Acesso em 04 nov. 2025. HAGIU, Andrei; WRIGHT, Julian. *Artificial intelligence and competition policy*. International Journal of Industrial Organization, 2025.

⁹ A Diretiva 790/2019 da União Europeia sobre direitos de autor no Mercado Único Digital define mineração de dados como “qualquer técnica de análise automática destinada à análise de textos e dados em formato digital, a fim de produzir informações, tais como padrões, tendências e correlações, entre outros” UNIÃO EUROPEIA, *Diretiva (UE) 2019/790 do Parlamento Europeu e do Conselho, de 17 de abril de 2019, relativa aos direitos de autor e direitos conexos no mercado único digital e que altera as Diretivas 96/9/CE e 2001/29/CE*, art. 2º, (2).

¹⁰ BENDER, E. M., GEBRU, T., MCMILLAN-MAJOR, A., & SHMITCHELL, S. (2021). *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?*. FAccT 2021.

Esse fator tem levado a diversas ações judiciais, principalmente nos Estados Unidos,¹¹ sede da maioria das empresas de IAGs que ganharam destaque nos últimos anos, relativas a diferentes modalidades, como texto (v.g. *New York Times v. Microsoft* e *OpenAI*¹²), áudio (v.g. ações de grandes gravadoras contra as desenvolvedoras de IAG Suno Inc.¹³ e Udio¹⁴) e imagens (v.g. *Andersen et al v. Stability AI et al*, ação coletiva promovida por artistas visuais em face das desenvolvedoras das ferramentas StableDiffusion, DreamStudio, Midjourney e DreamUp¹⁵). Também há disputas no Brasil, pretendendo indenização por uso não autorizado de obras autorais para treinamento de modelos de IA (na ação da Folha de S. Paulo contra a OpenAI¹⁶), ou por reprodução de músicas geradas por IAG em locais públicos (na ação declaratória de não infração promovida pela Spitz Park Aventuras Ltd. questionando cobranças do ECAD¹⁷), ou por direitos da personalidade como uso de voz alegada semelhante à do autor (como na ação promovida por dublador contra shopping center envolvendo campanha publicitária gerada com IAG¹⁸). Poucas ações chegaram a uma decisão e algumas resultaram em acordos, sem análise do mérito, além de haver acordos paralelos, independentemente de ação deduzida em juízo.¹⁹

¹¹ Vide, por exemplo, lista não exaustiva em: <https://chatgptiseatingtheworld.com/2025/01/03/updated-map-of-copyright-lawsuits-v-ai-companies/>. Acesso em 23 out. 2025.

¹² ESTADOS UNIDOS. *New York Times v. Microsoft and OpenAI*, Case No 1:23-cv-11195, U.S. District Court for the Southern District of New York.

¹³ ESTADOS UNIDOS. *UMG Recordings, Inc. et al. v. Suno Inc. et al.*, Case No 1:24-cv-11611, U.S. District Court for the District of Massachusetts. Disponível em: <https://www.pillsburylaw.com/a/web/mASCCJ5JhVmQH2dGyXYF77/9f8pZo/suno-complaint.pdf>. Acesso em 05/11/2025.

¹⁴ ESTADOS UNIDOS. *UMG Recordings Inc. et al. v. Uncharted Labs Inc., d/b/a Udio.com et al.*, Case No 1:24-cv-04777, U.S. District Court for the Southern District of New York. Disponível em: <https://www.pillsburylaw.com/a/web/sxgpKdHK22HVN4V7YxFcy8/9f8pZ8/udio-complaint.pdf>. Acesso em 05/11/2025.

¹⁵ ESTADOS UNIDOS. *Andersen v. Stability AI Ltd*, Case No 3:23-cv-00201, U.S. Northern District of California.

¹⁶ TJSP. 3ª Vara Empresarial e Conflitos de Arbitragem da Comarca de São Paulo/SP, Processo nº 1107237-96.2025.8.26.0100. Requerente: Empresa Folha da Manhã S/A. Requeridos: OpenAI Incs. OpenAI Global, LLC, OpenAI GP, LLC.

¹⁷ TJSC. 6ª Vara Cível da Comarca de Florianópolis/SC, Ação de Inexistência de Débito n. 5028302-65.2025.8.24.0023. Requerente: Spitz Park Aventuras Ltda. Requerido: Escritório Central De Arrecadação e Distribuição – ECAD.

¹⁸ TJSP. 18ª Vara Cível da Comarca de São Paulo/SP, Processo nº 1119021-41.2023.8.26.0100. Requerente: Igor Lott Zeger Belkind. Requerido: Shopping Anália Franco.

¹⁹ Os acordos realizados não implicam reconhecimento de direitos autorais envolvidos, como o acordo de US\$20 milhões anuais pagos pela Amazon ao New York Times (BRUELL, Alexandra. *Amazon to Pay New York Times at Least \$20 Million a Year in AI Deal*. Wall Street Journal. Julho 2025. Disponível em: <https://www.wsj.com/business/media/amazon-to-pay-new-york-times-at-least-20-million-a-year-in-ai-deal-66db8503>. Acesso em 9 out. 2025, ou o acordo entre OpenAI e Axel Springer na ordem de dezenas de milhões de euros BLOOMBERG. *OpenAI to Pay Axel Springer Tens of Millions to Use News Content*. <https://www.bloomberg.com/news/articles/2023-12-13/openai-axel-springer-ink-deal-to-use-news-content-in-chatgpt>. Acesso em 24 out. 2025).

A questão é particularmente relevante dentro do debate legislativo sobre inteligência artificial no Brasil, o PL 2.338/2023²⁰, bem como sobre a aplicação da Lei de Direitos Autorais (Lei n. 9.610, de 19 de fevereiro de 1998).²¹

O PL 2.338/2023 contém seção específica para tratar “Dos Direitos do Autor e Conexos” (Seção IV do Capítulo VIII – arts. 62 a 66), prevendo, dentre suas regras, a obrigação de os desenvolvedores de modelos de IAG: (i) indicarem especificamente as obras protegidas usadas na base de dados para o treinamento do modelo (art.62); (ii) gerenciarem o consentimento dos autores no uso das obras para desenvolvimento de IA (direito de objeção de uso pelos autores previsto no art. 64); e (iii) remunerarem os autores pelo uso de suas obras no treinamento de modelos de IA, especificando inclusive com condições de negociação coletiva, nos termos da Lei de Direitos Autorais (art. 65). Prevê-se exceção exclusivamente para a mineração de dados para fins de pesquisa e desenvolvimento de modelos de IA por determinadas organizações, sem fins comerciais (art. 63).

Ou seja, se não menciona explicitamente, o regramento praticamente pressupõe a existência de direito autoral envolvido no uso de obras protegidas para treinamento de modelos de IAG, prevendo a necessidade de autorização e remuneração correspondente.

1.2. Síntese do argumento

A Lei de Direitos Autorais prevê a obrigatoriedade de autorização para uso da obra protegida, nos seguintes termos e hipóteses:

*Art. 29. Depende de autorização prévia e expressa do autor a **utilização da obra**, por quaisquer modalidades, tais como:*

I - a reprodução parcial ou integral; (...)

IX - a inclusão em base de dados, o armazenamento em computador, a microfilmagem e as demais formas de arquivamento do gênero;

X - quaisquer outras modalidades de utilização existentes ou que venham a ser inventadas. (grifamos).

Pela abrangência que atribui ao conceito de uso da obra, esse dispositivo é por vezes mencionado como “previsão explícita” da necessidade de autorização para utilização da obra

²⁰ Versão aprovada pelo Senado e encaminhada à Câmara dos Deputados, onde tramita atualmente. Disponível em: https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra?codteor=2868197&filename=PL%202338/2023. Acesso em 23 out. 2025.

²¹ PL 2338/2023, Art. 65.

para treinamento de modelos de IAG, citando-se em particular o inciso IX que trata da *inclusão da obra em bases de dados*.

Importante notar que o treinamento de modelos IAG pode ser realizado sem a cópia ou transferência direta dos dados de treinamento, por meio de técnicas como *federated learning* e *split learning*, nas quais os dados permanecem localmente nas instituições ou dispositivos que os geraram, sendo apenas os parâmetros atualizados compartilhados com o modelo central. Portanto, a inclusão ou não em base de dados não pode ser o ponto decisivo para a proteção autoral. O ponto chave diz respeito ao tipo de processamento realizado pela máquina, se voltado ou não para projetar a obra em sua expressão individual para o entendimento humano.

O que parece passar despercebido é que o art. 29, quando fala em qualquer utilização, “*existente ou que venha a ser inventada*”, como objeto de proteção, obviamente fala de qualquer utilização *da obra como tal*, isto é o uso dos dados correspondentes para expressão do seu sentido individualizado na comunicação entre autor e público. Se, no campo analógico, o uso dos dados necessariamente é uso expressivo da obra, pois a compreensão humana é imediata, no campo digital, a compreensão humana é mediatizada pela máquina e há uma série de uso de dados codificados que não se destinam à projeção ou expressão da obra individual. Processamentos computacionais diversos como compressão de arquivos, cópia de segurança, verificação de *malwares*, integridade ou exfiltração de dados, podem implicar cópia, transferência, armazenamento ou inclusão em base de dados, sem que se cogite de proteção autoral.

Isso porque o objeto de proteção do direito autoral, a obra intelectual artística ou literária, é a expressão individual de uma ideia pelo autor, materializada em determinado suporte. Essa *expressão individualizada* forma o conteúdo de uma comunicação entre autor e público, por meio da obra, comunicação esta que pressupõe a possibilidade de apreensão de seu sentido. Esse pressuposto, porém, não está presente na utilização de dados digitais derivados de obras artísticas ou literárias para treinamento de sistemas de IAG. *Primeiro*, porque na digitalização para processamento computacional, os dados são traduzidos em determinada representação matemática e seu processamento envolve apenas a manipulação sintática de números, sem expressão de sentido ou possibilidade de comunicação entre autor e público. *Segundo*, porque o resultado do processamento lógico no treinamento de IAG de propósito geral é um modelo estatístico de representação do agregado de dados digitalizados, correspondentes a obras, que capta padrões gerais, conceitos, estilos, traduzidos por relações entre atributos com pesos relativos (parâmetros), que não reproduzem, nem armazenam aspectos individuais de cada obra digitalizada usada no treinamento. Na construção do modelo de IAG, assim, não

há possibilidade de comunicação de sentido, nem é armazenada ou memorizada a representação digital de qualquer obra individual.

Daí porque, na análise desenvolvida na **Seção 2**, concluímos **inexistir utilização expressiva da obra individual no treinamento de IAG de propósito geral, não se aplicando a proteção autoral**. Não há cópia, reprodução, armazenamento ou qualquer forma de utilização da obra (sua expressão individual). Apenas os dados digitais, dos quais as obras individuais poderiam ser projetadas, são utilizados para outra finalidade, são insumos em uma espécie de “moinho de dados” que resultará na produção de um modelo geral de IAG que, em si, não projeta, nem guarda ou armazena qualquer obra individual.

Respondida negativamente à questão jurídica sobre a hipótese de violação a direito de autor pelo uso de obras para treinamento de modelos de IAG de propósito geral, pode-se questionar se, em sede de deliberação legislativa, seria *justo* ou *politicamente recomendável* remunerar o uso desses “insumos digitais”, que advém, em última análise, do processo criativo de autores. Sem trazer posicionamento sobre essa questão, mostraremos, na **Seção 3**, ser inadequado e indesejável basear essa remuneração nos moldes do direito autoral, tal como proposto no PL 2338/23, ou seja baseada na exigência aos desenvolvedores de modelos de IAG (i) de identificação individualizada das obras protegidas contidas nos dados usados para treinamento; (ii) de gestão de consentimento pelos autores (iii) de remuneração individualizada aos autores das obras.

Tais obrigações elevam demasiadamente os custos de desenvolvimento dos modelos de IA, antes que os mesmos possam sequer encontrar aplicação ou exploração comercial no mercado. Os mercados de IA vivem momento de experimentação e desenvolvimento tecnológico e sua oneração excessiva pode favorecer grandes empresas líderes em detrimento de entrantes e empresas de pequeno porte, levando a potencial concentração de mercado e limitação ao desenvolvimento do Brasil nesse setor. Por outro lado, tal modelo de remuneração pode acabar alijando a produção cultural e artística nacional dessa nova tecnologia, cada vez mais popular entre cidadãos brasileiros, levando a uma forma certamente indesejável de colonialismo digital.

1.3. Especificação do objeto deste estudo

Como se observa pela síntese acima, este estudo limita-se a indagar sobre violação autoral pelo uso de obras para treinamento de modelos de IAG. O debate em torno dessa questão, porém, por vezes, em particular nos tribunais, confunde-se com questões laterais relativas a outras etapas do ciclo de vida de IA. Assim, antes de desenvolvermos o argumento para responder a essa questão central, vale esclarecer, a partir de debates travados em tribunais de

diferentes jurisdições, aquilo que não fará parte deste estudo.

Vale começar pelo caso *Thomson Reuters v. Ross Intelligence*.²² Naquele caso, decidiu-se que constitui violação de direitos autorais o uso desautorizado de memorandos jurídicos da Reuters para desenvolver a IA Ross, o qual não seria considerado uso transformativo capaz de fazer incidir a exceção de *fair use*.²³ A IA Ross, porém, não é generativa, mas uma IA preditiva treinada especificamente para, a partir de briefings jurídicos, predizer trechos de decisões que estejam relacionadas ou respondam a questões presentes no briefing apresentado. Assim, diferentemente de IAGs, que são de propósito geral (ver caracterização no Anexo Técnico, seção 5), a Ross Intelligence usou conteúdo da Thompson Reuters e outras bases de precedentes, como a WestLaw, para treinar a IA Ross a destacar trechos daqueles conteúdos que provavelmente seriam boas respostas às questões colocadas, sendo, portanto, usada para a mesma finalidade dos documentos comercializados pela autora. Entretanto, **o treinamento de modelos de IA para representar determinada obra ou autor específico não será objeto deste estudo.**

Outro caso importante para o esclarecimento do objeto a ser tratado aqui é *Bartz v. Anthropic*.²⁴ Naquele caso, o tribunal entendeu ser justo o uso de livros protegidos para treinar seu modelo Claude, mesmo sem autorização pelos autores dos livros digitalizados que compunham a base de dados de treinamento. No entanto, o tribunal decidiu haver infração pelo acesso não autorizado ao conteúdo digital, que foram baixados em sites considerados “piratas”. O caso terminou em acordo com pagamento de compensação de US\$1,5 bilhões pela Anthropic para resolver as alegações de uso de livros pirateados.²⁵ Não obstante, **a licitude do acesso a conteúdo autoral disponibilizado na internet não será objeto deste estudo.**

Os casos *NY times versus OpenAI* e *Folha de São Paulo versus OpenAI* também trazem aspectos laterais que não serão tratados aqui. Naqueles casos, os autores trouxeram exemplos de prompts ou requisições na fase de inferência ou emprego do *ChatGPT* que geravam resultados semelhantes ou muito próximos, às vezes com trechos literais, de conteúdo de propriedade dos autores. Por vezes, o conteúdo gerado nas ações foi obtido com *upload* de documentos autorais para que então o ChatGPT fosse instado produzir algo semelhante ou

²² Thomson Reuters v. Ross Intelligence, Case No 1:20-cv-613-SB, District Court for the District of Delaware.

²³ BAUTISTA, Lucia. *Delaware District Court Rules No Fair Use in Artificial Intelligence Case*. The Columbia Journal of Law & the Arts. August 17, 2025. Disponível em: <https://journals.library.columbia.edu/index.php/lawandarts/announcement/view/816>. Acesso em 20 out. 2025.

²⁴ Bartz v. Anthropic, Case No 3:24-cv-05417, Northern District of California.

²⁵ BAKER HOSTETLER. Bartz v. Anthropic. Disponível em: <https://www.bakerlaw.com/bartz-v-anthropic/>. Acesso em 20 out. 2025.

no estilo do colunista ou jornalista. **A licitude do emprego de sistemas de IAG pelo usuário para gerar de conteúdo similar a obra ou estilo de determinado autor não será objeto deste estudo.**

Também no caso *Spitz versus ECAD*, o Tribunal de Santa Catarina entendeu em decisão liminar de agravo, que deveria haver pagamento ao ECAD por execução de músicas geradas pelo aplicativo Suno, em parque de diversões. **A responsabilidade ou dever de remuneração autoral por conteúdo gerado ou uso de conteúdo gerado por IAG não será objeto deste estudo.**

Como se pode ver, há questões bem distintas, por vezes confundidas, nos temas levados aos tribunais naquelas ações judiciais. Uma delas é a questão sobre a licitude do acesso a conteúdo. Outra é a questão sobre a licitude do uso de obras autorais para treinamento de *modelos* de IAGs. Também é diferente a questão sobre a licitude da especificação ou ajuste fino (*fine-tuning*) de um modelo geral de IAG para determinada obra ou produção de determinado autor. Por fim, outra questão independente é a licitude do emprego de *sistemas* de IAG para, na fase de inferência, provocar e direcionar na requisição (ou *prompt*)²⁶ a geração de conteúdo que seja similar a determinada obra ou que use o estilo de determinado autor.²⁷

Vale breve incursão nas diferentes etapas no ciclo da IAG, desde a coleta de dados, treinamento até a inferência, para esclarecer essas distinções.

(1) Acesso ao Conteúdo

O ponto de partida do ciclo de desenvolvimento de sistemas de inteligência artificial (IA) é o acesso ao conjunto de dados que servirá de base para o treinamento. Esses dados podem incluir textos, imagens, vídeos, áudios e outros materiais. O aspecto que pode ser analisado nesta etapa é a legitimidade do acesso, ou seja, se os dados estão disponíveis publicamente ou submetidos a restrições contratuais, técnicas ou legais. Isso abrange questões como a publicidade dos dados coletados, a existência de barreiras técnicas de autenticação ou *paywalls*, a ausência de autorização expressa e o uso de conteúdos pirateados²⁸. O acesso pode ser ilícito caso desrespeite os condicionamentos colocados pelo autor ou por aquele que disponibiliza a obra na *internet*.

²⁶ Instruções que combinam um contexto (informações, exemplos ou arquivos fornecidos pelo usuário) e uma requisição (o comando que orienta a geração de uma resposta).

²⁷ WOLFSON, Stephen. *The Complex World of Style, Copyright, and Generative AI*. Creative Commons. Disponível em: <https://creativecommons.org/2023/03/23/the-complex-world-of-style-copyright-and-generative-ai/>. Acesso em 27 out. 2025.

²⁸ Como discutido no caso norte-americano *Bartz v. Anthropic*, analisado mais adiante.

(2) Treinamento do modelo de propósito geral de IAG

Superada a etapa de coleta, os dados são empregados no treinamento do modelo de IAG (também denominado modelo fundacional ou modelo de propósito geral). Esse processo consiste em submeter o sistema a grandes volumes de dados para que ele aprenda por métodos estatísticos padrões de escrita, voz, composição musical, imagens e vídeos, para que o sistema desenvolvido a partir do modelo seja capaz de gerar conteúdos inéditos. O processo de desenvolvimento desses modelos é descrito em detalhe no ANEXO TÉCNICO e, como se pode ver, o treinamento resulta em um modelo de representação que não retém ou se refere a obras individuais, mas a abstrações e atributos conceituais ligados a parâmetros (ou pesos que indicam a importância relativa dos conceitos e atributos) extraídos do agregado de obras.

(3) *Fine-tuning*

Após o treinamento do modelo de propósito geral, pode-se realizar *fine-tuning* (ajuste fino), uma etapa complementar destinada a especializar o modelo em determinado domínio ou *corpus* específico de conteúdo.²⁹ Trata-se de processo de refinamento, no qual o modelo fundacional é novamente exposto a um subconjunto de dados, com o objetivo de ajustar seus parâmetros e torná-lo mais sensível a contextos, linguagens ou tarefas particulares.

O *fine-tuning* representa, portanto, uma transição entre o aprendizado genérico de um modelo de propósito geral e a especificação do modelo para que o sistema correspondente gere resultados mais apropriados para determinado fim. O refinamento pode ocorrer em diferentes níveis de generalidade/especificidade. Por exemplo, refinar um modelo de IAG para músicas em geral, direcionando para música brasileira, ou para o estilo Bossa Nova ou especificamente para o estilo das composições ou voz de João Gilberto. O Projeto *The Next Rembrandt* (<https://www.vml.com/work/next-rembrandt>) é um exemplo de IA – que não chega a empregar IA generativa, mas modelo de *deep learning* e reconhecimento facial (ver Anexo Técnico, item 4.3. para entender os modelos de deep learning e reconhecimento facial) – voltada para autor específico, Rembrandt, com base em obras de domínio público.

Assim, a depender do grau de especificidade, particularmente, quando o treinamento se dirige a associações reconhecíveis a um autor, estilo ou obra específica com a finalidade de mimese pelo sistema de IAG resultante, pode-se indagar se haveria violação autoral, ou ainda, se o emprego do sistema de IAG resultante do *fine-tuning* geraria conteúdo violador a direito autoral.

²⁹ XU, L., XIE, H. & QIN, S-Z. J. et al. (2023). *Parameter-Efficient Fine-Tuning Methods for Pretrained Language Models: A Critical Review and Assessment*. arXiv preprint arXiv:2312.12148. Disponível em <https://arxiv.org/abs/2312.12148> (acesso em 05.11.2025)

(4) Aplicação de Sistema de IAG para geração de conteúdo

A etapa final do ciclo é a aplicação de um sistema de IAG programado a partir de um modelo fundacional, isto é, o momento em que o sistema é utilizado por meio de requisições ou *prompts* para gerar resultados (*outputs*), textuais, visuais ou sonoras.

O prompt pode ser direcionado, em seu contexto ou requisição a determinada obra ou autor, solicitando-se seu conteúdo integral ou parcial ou ainda alguma variação deste. Há situações (como nas ações NYT versus OpenAI e FSP vs OpenAI) em que o próprio usuário insere no *prompt* trechos, imagens ou áudios de exemplo (*sample inputs*) e pede que o modelo os reproduza ou modifique.

Há ainda a possibilidade de uso de RAG (Retrieval-Augmented Generation), que consiste em combinar um modelo de IAG com um mecanismo de busca de informações externas que são recuperadas em tempo real e integradas ao processo de geração.³⁰ Em vez de depender apenas do conhecimento adquirido no treinamento do modelo, o RAG permite que o sistema consulte bases de dados específicas — como documentos jurídicos, artigos científicos ou manuais técnicos — para produzir respostas mais contextualizadas e atualizadas sobre determinado domínio. Essa técnica difere do *fine-tuning*, que envolve ajustar os pesos internos do modelo a partir de novos dados, criando uma versão especializada, porém estática. O RAG, por sua vez, não altera o modelo original, mas amplia seu alcance informativo por meio da recuperação dinâmica de conteúdo relevante para as especificações feitas pelo usuário já na fase de inferência do sistema de IAG desenvolvido a partir do modelo.

Outro ponto, tratado mais adiante e especificado no Anexo Técnico (item 6), mas que também se refere à fase de inferência, diz respeito à possibilidade técnica de “memorização” pelo modelo ou “regurgitação” pelo sistema de IAG de trechos próximos a obras usadas no treinamento.

O quadro abaixo sintetiza as questões pertinentes a cada etapa do ciclo de vida da IA.

(1) Acesso a conteúdo	(2) Treinamento	(3) <i>Fine-tuning</i>	(4) Uso do sistema (inferência)
O acesso às obras para formação de bases de dados foi lícita (obra não estava disponível ao público ou havia	Haveria violação autoral pelo treinamento do modelo de IAG a partir de obras autorais, sem	Haveria violação autoral pelo refinamento de modelo geral de IAG com base em	Haveria violação autoral quando o usuário faz requisições para direcionar o sistema e provocar a

³⁰ Para uma introdução à técnica de RAG, ver GUPTA, Shailja; RANJAN, Rajesh; SINGH, Surya Narayan. A comprehensive survey of retrieval-augmented generation (rag): Evolution, current landscape and future directions. arXiv preprint arXiv:2410.12837, 2024.) Disponível em <https://arxiv.org/abs/2410.12837> (acesso 05.11.2025))

condicionamentos para acesso)?	autorização pelos respectivos autores?	conjunto de obras de determinado autor, para que o sistema gere conteúdos similares a obras ou no “estilo” de determinado autor?	geração de conteúdo similar ou com trechos idênticos a obra protegida presente na base de treinamento?
--------------------------------	--	--	--

Quadro 1. Etapas do ciclo de sistemas de IAG e questões envolvidas. Elaboração própria.

Com as distinções feitas acima, podemos evitar confusões que têm, de certa forma, gerado ruídos no debate público sobre IAGs e direitos autorais, e especificar o objeto deste estudo. Trataremos apenas da **questão 2** indicada no quadro acima. Ou seja, **este estudo analisa unicamente a seguinte questão jurídica: há violação autoral pelo uso, sem autorização, de obras protegidas na base de dados empregada para treinamento de modelos de IAG?** As considerações aqui presentes não autorizam conclusões sobre acesso a obras disponíveis na internet, sobre direitos autorais relativos a *fine-tuning* do modelo direcionado a obra ou autores específicos, ou sobre direitos autorais decorrentes de conteúdo gerado por sistema de IAG a partir de contextualização ou requisição direcionada pelo usuário a determinado autor ou obra.

2. A OBRA AUTORAL E SUA PROTEÇÃO

Inovações tecnológicas transformadoras usualmente trazem desafios a conceitos fundamentais do direito de propriedade intelectual.³¹ Não é diferente com a IAG, que talvez leve ao limite o esgarçamento do direito autoral, tal como concebido na era moderna,³² ao impor questões profundas sobre o próprio conceito de *criação* e *autoria*, pela geração de conteúdos inéditos por máquina, e sobre a própria noção de *obra* e de *utilização da obra* (cópia, reprodução ou armazenamento) que será particularmente explorado neste estudo. Desafios conceituais deste porte, impõem à exegese indagar sobre o propósito e fundamentos jurídicos do próprio instituto.³³

2.1. A obra

O direito autoral, com a feição de um direito subjetivo e de personalidade do autor, nasce de uma construção jurídica e ética, bastante abstrata, na França e Estados Unidos, no contexto das revoluções iluministas, para substituir o então desgastado modelo de privilégio estatal conferido a editores/impressores para reproduzir cópias de obras com exclusividade (de onde o termo *copyright*).³⁴ Nessa moldura de pensamento moderno, as justificativas éticas aparecem centradas na figura autor, como um direito subjetivo a ser assegurado pelo e defendido perante o Estado. O desafio estava em fundamentar moralmente, perante a noção de autonomia individual, como um bem não-rival (o desfrute de uma obra por um, não exclui o desfrute pelo outro), poderia ser tratado como uma espécie de “propriedade física”. A solução foi se concentrar no aspecto de materialização da obra, revestindo-a da prerrogativa central do direito de propriedade, que é o uso exclusivo do bem. A autonomia do autor seria protegida pelo seu direito exclusivo de autorizar reproduções materiais da própria criação intelectual.

Na tradição anglo-saxã, a construção é pragmática, primeiro com a justificativa lockeana baseada na remuneração ao esforço ou trabalho, mas ligado à individualidade. A propriedade é legitimamente adquirida ao se estender a propriedade sobre a própria identidade mental e extensão corpórea (daí porque Locke inicialmente usava o termo *propriety* e não *property* para se referir à propriedade), empreendendo esforço, no caso, o esforço criativo, que agrega

³¹ GOLDSTEIN, Paul. *Copyright's Highway: From the Printing Press to the Cloud*. Stanford University Press, 2003.

³² MARANHÃO, J. Inferências metafóricas e a reconfiguração do Direito na era digital. *Revista de Direito e das Novas Tecnologias*, v. 1, ano 1, p. 15-30, 2018.

³³ BARAK, Aharon. *Purposive interpretation in law*. Princeton: Princeton University Press, 2011.

³⁴ GINSBURG, Anne. A tale of two copyrights: literary property in revolutionary France and America. *Tulane Law Review*, v. 64, p. 993 a 1031, 1990.

valor à sociedade.³⁵ Por agregar valor, é legítima a apropriação de uma porção desse ganho social. Essa justificativa da propriedade intelectual pela justa remuneração ao esforço individual, como direito subjetivo, evolui, com Hume³⁶ e diferentes versões utilitaristas subsequentes, para uma justificativa social:³⁷ para que haja incentivo à produção intelectual, industrial ou cultural, o custo de produção precisa ser recompensado, mas não além do necessário para induzir o autor à essa troca social.³⁸ Essa medida vai ser calibrada, diante da crescente redução dos custos de reprodução por terceiros como avanço tecnológico e a preocupação com o oportunismo na exploração das reproduções.³⁹

Na construção continental, a justificativa, de raiz kantiana e hegeliana, assenta-se em direito absoluto, na medida em que o ser humano, dotado de autonomia, tem na expressão de uma criação intelectual, por sua vontade, uma forma de desenvolvimento e realização da própria personalidade.⁴⁰ Assim, na construção kantiana, a proteção da autoria é uma dimensão da própria dignidade humana. Na matriz hegeliana, o fundamento ético está não no ganho social que justifica a remuneração pelo esforço empreendido, mas na proteção da “persona” (a imagem, honra e história individual), que expressa e forma sua personalidade na esfera pública, por meio da obra.⁴¹

Tais diferenças no fundamento ético explicam nuances na proteção resultante, como a maior ênfase continental nos direitos morais ou a maior flexibilidade no direito anglo-saxão para reconhecer direitos quando um produto traz ganhos sociais ainda que não seja clara a contribuição pessoal na criação, ou para não os reconhecer quando o uso por terceiros traz ganhos sociais sem prejudicar a exploração econômica do autor original (a doutrina norte-americana do *fair use*). Essas nuances são importantes para entender a forma como vem sendo abordado o treinamento de modelos de IAG no direito europeu e anglo-saxão.

Mas em ambas as abordagens, o objeto da proteção autoral está na expressão individual de uma criação em determinada extensão material trazida a público, seja pela visão

³⁵ LOCKE, John. Segundo tratado sobre o governo. Tradução de Magda Lopes e Marisa Lobo da Costa. Petrópolis: Vozes, 1994.

³⁶ HUME, David. Tratado da natureza humana. Tradução de Déborah Danowski. São Paulo: Editora UNESP, 2009.

³⁷ Essa visão orientada à sociedade, em oposição à concepção orientada ao autor, típica da construção continental, já aparece na cláusula de direito autoral da Constituição norte-americana, ecoando o Statute of Anne, inglês (ver GINSBURG, op. cit.)

³⁸ LANDES, William M.; POSNER, Richard A. *An Economic Analysis of Copyright Law*. Journal of Legal Studies, v. 18, n. 2, p. 325–363, 1989.

³⁹ BARBOSA, Denis Borges. *Direito de Autor. Questões Fundamentais de Direito de Autor*. 1ª ed., Lumen Juris, 2013, p. 66; HUGHES, Justin. *The Philosophy of Intellectual Property*. Georgetown Law Journal Association, 1988; STANFORD ENCYCLOPEDIA OF PHILOSOPHY. Intellectual Property. 2022. Disponível em: <https://plato.stanford.edu/entries/intellectual-property/#JustCrit>. Acesso em 23 out. 2025.

⁴⁰ MERGES, Robert P. *Justifying Intellectual Property*. Cambridge, MA: Harvard University Press, 2011.

⁴¹ FISHER, William W. "Theories of Intellectual Property." In *New Essays in the Legal and Political Theory of Property*, ed. Stephen Munzer. Cambridge, UK: Cambridge University Press, 2001.

continental deontológica de que a obra é uma extensão da própria personalidade individual do autor (*authorship norm*), seja pela ou justificativa consequencialista de incentivo à cultura pela compensação ao autor pelo custo específico de determinada produção intelectual (*marketplace norm*). Ideias, temas, ou conceitos, por sua vez, não são protegidos, seja por não consubstanciarem a *persona* individual, que só aparece nessa expressão material a ser comunicada ao público, seja para não limitar a produção cultural.⁴²

Esses traços estão bem-marcados no direito brasileiro, de matriz continental. Assim, a Lei de Direitos Autorais define como objeto de proteção, “*criações do espírito, expressas por qualquer meio ou fixadas em qualquer suporte*”⁴³, excluindo desse objeto as ideias que tais obras materializam (art. 8º, inc. I) ⁴⁴, com base na concepção de que, como coloca COSTA NETTO, “*cada criador tem um modo distinto de decodificar a ideia, ou seja, quando a materializa o faz colocando suas características pessoais. E é exatamente o resultado materializado desta ‘decodificação’ que o direito de autor protege.*”⁴⁵ Daí, como coloca OLIVEIRA ASCENSÃO, se extrai a noção de *individualidade* como elemento característico decisivo para que se fale em obra.⁴⁶

A obra, assim, não é a matéria que lhe dá suporte físico, nem a ideia que é consubstanciada nessa expressão individual por determinado autor. Ela é o sentido específico e individual resultado daquele ato expressivo, mas que se constrói e é percebida na interação entre autor e público. Antônio Cândido a concebe como uma relação dialética entre a expressão material, autor e público.

*“Na medida em que arte é –como foi apresentada aqui– um sistema simbólico de comunicação inter-humana, ela pressupõe o jogo permanente de relações entre os três, que formam uma tríade indissolúvel. O público dá sentido e realidade à obra, e sem ele o autor não se realiza, pois ele é de certo modo o espelho que reflete a sua imagem enquanto criador.... Deste modo, o público é fator de ligação entre o autor e sua própria obra. A obra, por sua vez, vincula o autor ao público, pois o interesse deste é inicialmente por ela, só se estendendo à personalidade que a produziu depois de estabelecido aquele contato indispensável. Assim, à série autor-público-obra, junta-se outra: autor-obra-público. Mas o autor, do seu lado, é intermediário entre a obra, que criou, e o público, a que se dirige; é o agente que desencadeia o processo, definindo uma terceira série interativa: obra-autor-público.”*⁴⁷

A obra é em uma relação de comunicação entre autor e público e somente a partir dessa

⁴² SAMUELS, E. The idea-expression dichotomy in copyright law. DRASSINOWER A. A Rights-Based View of the Idea/ Expression Dichotomy in Copyright Law. *Canadian Journal of Law & Jurisprudence*. 2003;16(1):3-21. doi:10.1017/S0841820900006603. Para uma visão crítica da distinção ver JONES, R. The Myth of the Idea/Expression Dichotomy in Copyright Law. 10 Pace Law Rev. 551 (1990).

⁴³ Lei 9.610/1998, art. 7º, caput.

⁴⁴ Lei 9.610/1998. “Art. 8º Não são objeto de proteção como direitos autorais de que trata esta Lei: I - as idéias [...]”.

⁴⁵ COSTA NETTO, José Carlos. *Direito Autoral no Brasil*. 5. ed. São Paulo: SaraivaJur, 2025.

⁴⁶ ASCENSÃO, José de Oliveira. *Direito Autoral*. 3. ed. Rio de Janeiro: Renovar, 2007, p. 35-36.

⁴⁷ CÂNDIDO, Antônio. *Literatura e Sociedade*. Editora Nacional, São Paulo, 1965, pp. 44-45.

relação comunicativa é possível concebê-la. Portanto, outro pressuposto fundamental para que se fale em “obra” como obra intelectual é a sua capacidade de sentido, ou seja, como destaca OLIVEIRA ASCENSÃO, “*tem de se exteriorizar ou manifestar por meio que seja captável pelos sentidos*”.⁴⁸

2.2 A reprodução ou utilização da obra: analógica versus digital

O entendimento do conceito de “reprodução”, “cópia”, “armazenamento” ou “qualquer forma de utilização” é corolário da noção de “obra” como materialização, dotada de sentido, da expressão individual do artista. O uso da obra, como tal, em qualquer de suas acepções ou modalidades, em qualquer mídia ou tecnologia, deve envolver alguma forma ou consistir em parte de processo que resulte na comunicação autor-público. O objeto de proteção, na verdade, é essa comunicação, que, na visão continental, projeta na esfera pública a personalidade individual do autor, consubstanciada na obra, e que, na visão anglo-saxã, promove a cultura e justifica a remuneração do esforço. Sua reprodução, cópia ou utilização, para que seja uma *cópia ou utilização da obra*, capaz de afetar o direito do autor, sua imagem, honra e sua exploração econômica também pressupõe a capacidade de sentido dessa utilização ou possibilidade de inteligibilidade pelo humano.

Esta última consideração parece óbvia quando tratamos de obras materializadas ou reproduções e cópias em suporte analógico, em que a apreensão de sentido é imediata. Mas há nuances quando tratamos de uma comunicação intermediada pela digitalização. Entender esses nuances é crucial para interpretar o art. 29 da Lei de Direitos autorais, quando exige a autorização prévia do autor para a “*utilização da obra, por quaisquer modalidades*”, como “*a reprodução parcial ou integral*”, a “*inclusão em base de dados, o armazenamento em computador*” ou ainda “*quaisquer outras modalidades de utilização (...) que venham a ser inventadas*”.

Para tanto, é preciso distinguir e compreender a relação entre “*dados*” e a “*informação*” deles extraída, que é o conteúdo de sentido apreendido pelo humano.

De modo abstrato, dados podem ser definidos como quebras de uniformidade perceptíveis (pelo humano ou por máquina)⁴⁹. A marca de tinta azul em um papel branco é percebida pela visão humana como uma quebra de continuidade e, ao formar um símbolo em determinada linguagem ou gramática, é percebido pelo humano como sinal distintivo. Os bits, sinais elétricos, por sua vez, são percebidos pelo processamento computacional

⁴⁸ ASCENSÃO, José de Oliveira. *Direito Autoral*. 3. ed. Rio de Janeiro: Renovar, 2007, p. 32-33.

⁴⁹ FLORIDI, Luciano. *The Logic of Information: A Theory of Philosophy as Conceptual Design*. Oxford: Oxford University Press, 2019.

(presença/ausência do sinal).

Na percepção humana, a detecção dos dados analógicos e a extração de significado é *imediata*. Quando um humano competente na linguagem identifica uma sequência de letras organizada segundo determinada gramática, apreende o sentido das. Quando percebe auditivamente descontinuidades em amplitudes de ondas sonoras ou vibrações, capta sons, vozes etc. o que lhe traz entendimento ou emoções. Quando está diante de obra visual, como pintura, escultura ou instalação, percebe imediatamente cores, formas, relevos, profundidade e perspectivas que permitem a compreensão e despertam sensações. Também o humano percebe imediatamente, ao visualizar interfaces com cores, formas e movimentos, aliados a narrativas e músicas, o significado de obras audiovisuais.

Por sua vez, a percepção e compreensão de textos, áudios, imagens e vídeos *digitalizados* é intermediada pela máquina, fazendo com que a comunicação entre autor e público, pela obra nativo-digital ou digitalizada e reproduzida em meio digital, seja *mediata*. Os conteúdos analógicos são codificados e podem passar por diferentes formas de *processamento lógico*, para então serem decodificados e projetados para a percepção humana. O computador não entende, não enxerga não ouve e não compreende a continuidade do movimento. Basicamente, o computador representa o mundo em números e processa os processos sintaticamente, ao passo que o hardware, aliado a equipamentos, converte números em sinais físicos (luz, som, vibração) e só então o cérebro humano percebe os sinais físicos no mundo, compreendendo o significado (cognitivo ou emotivo), ao articulá-los (Ver no Anexo Técnico, item 1, a descrição do processo de codificação ou “*embedding*” de diferentes tipos de conteúdo).

Assim, qualquer reprodução ou utilização analógica imediatamente estabelece a relação autor-público, pois sua materialização no suporte é necessariamente dotada de sentido para o humano. Já entre a codificação em *bits* e a sua decodificação, o processamento lógico consiste em uma série de manipulações sintáticas, em representações numéricas binárias, que são, em oposição, incompreensíveis para os humanos. E quando uma obra, expressa em meio analógico, é digitalizada ou é produzida digitalmente, os dados digitais correspondentes podem ser processados para promover a projeção da obra em sua individualidade para percepção humana (decodificação da obra individual), mas podem também ser objeto de outras formas de processamento lógico que não resultarão na decodificação e projeção da obra individual.

Por exemplo, a obra pode ser digitalizada para ser disponibilizada na rede mundial de computadores, ou para transmiti-la a máquinas capazes de impressão com elevada precisão. Ou os dados digitais dos quais as obras são projeção podem ser incluídos em bases de dados, estruturada para organizar e sistematizar o acesso de determinada forma às obras

decodificadas, ou armazenadas em computadores com o fim de sua projeção posterior. Em todos esses casos, temos o tratamento dos dados no sentido de utilização da obra em sua individualidade, com etapas de um processo que resultará ou tem por finalidade resultar na comunicação autor-público do sentido daquela obra.

Porém, dados digitais correspondentes a obras artísticas e literárias, já armazenados em um servidor, podem passar por processos computacionais diversos, tais como a “compressão” para maior eficiência em seu armazenamento, a realização de cópia de segurança, para resgate dos dados em caso de exfiltração ou corrompimento, a transferência para uma base de dados de terceiros, em ambiente protegido para a verificação de presença de *malwares* ou vírus, ou reestruturação da base aprimorar controle de acessos não autorizados (serviços de segurança ou recuperação de dados). Nessas hipóteses, os dados digitais correspondentes à obra são manipulados sem o propósito de instaurar alguma comunicação autor-público, razão pela qual não se deve cogitar da aplicação de direitos autorais, ou exercício da prerrogativa de autorização prévia para inclusão em bases de dados, cópia ou armazenamento. Note-se, por exemplo, que a Lei de Software (Lei 9609/1998) excepciona explicitamente, em seu art. 6º, inc. I, a cópia de segurança.

O treinamento de modelos de IAG em si, como veremos a seguir, deve ser também visto como uma das formas de processamento lógico de dados digitais correspondentes a obras, mas que não envolvem ou não tem por propósito a projeção da obra para sua apreensão individualizada pelo humano (não instaura, nem tem por propósito instaurar a relação de comunicação autor-público), uma vez que tais modelos matemáticos (falamos aqui dos modelos apenas e não dos sistemas de IAG que fazem inferências a partir dos modelos, gerando conteúdos) limitam-se a extrair padrões que representam conceitos, ideias, estilos e temas a partir do agregado dos dados, sem armazenar ou memorizar obras em sua individualidade.

2.3. O treinamento de modelos de IAG não implica uso expressivo da obra individual

Dentre as diversas ações interpostas por autores ou detentores de direitos autorais contra desenvolvedores de modelos de IA, poucas chegaram a um pronunciamento judicial, pois ou pendem de decisão, dada a complexidade do novo tema levado aos tribunais, ou resultaram em acordo que levou a remuneração a autores, mas sem o reconhecimento de violação ou exercício de direito autoral.

Decisões por tribunais norte-americanos têm duas peculiaridades. Primeiro, o direito autoral tem matriz utilitarista e análise é construída sobre a doutrina do “fair use” em que se verifica se o uso consiste em um novo tipo transformador de criação (o que se coaduna ao

propósito de estímulo à produção cultural) e se tal criação traz concorrência que limita a exploração econômica do seu trabalho pelo autor (proteger a relação custo-benefício de produção cultural). Em segundo lugar, trata-se de sistema de *common law*, no qual as decisões judiciais são fonte de direito válido (criam direito). Assim, a separação entre uma *limitação ao exercício* de direitos de autor uma *exceção* à aplicação do direito de autor é mais tênue, pois a proteção jurídica consiste em atender à finalidade do instituto e aquilo que atende ou não a sua finalidade é determinada pela evolução da jurisprudência. Casos autorais inevitavelmente evocam a análise da doutrina do *fair use*, sem se questionar se efetivamente tratamos de aplicabilidade de direito de autor. Dois casos resultaram recentemente a pronunciamento judicial.

No caso *Bartz v. Anthropic* reconheceu-se que o treinamento do modelo de linguagem Claude é uso transformativo pois resulta em produto que cria conteúdos e não em cópias integrais ou parciais dos livros originais. Por outro lado, entendeu-se que o treinamento não acarreta competição com a exploração econômica das obras originais pelos autores. Alsup faz analogia entre o treinamento de IAG e o ensino a humanos: *"Authors' complaint is no different than it would be if they complained that training schoolchildren to write well would result in an explosion of competing works. This is not the kind of competitive or creative displacement that concerns the Copyright Act"*.⁵⁰

No caso de autores contra a *Meta*, o juiz Vince Chhabria da *Northern District of California Court*, apesar de reconhecer o uso transformativo, deu mais atenção à relação de competição da criação resultante com a exploração econômica da obra original pelo autor. Embora tenha admitido que modelos de IA podem inundar o mercado com obras que em geral podem exercer competição com a produção humana, entendeu que deveria haver demonstração de concorrência e limitação à exploração econômica de *obra específica*, o que não encontrou no caso sob sua análise, decidindo, portanto, pela ausência de violação autoral. Ou seja, sobrelevou-se naquela decisão, a nota da *individualidade* da obra, que é o objeto de proteção autoral.

Note-se que a doutrina de *fair use* acaba por abordar etapas distintas do processo computacional envolvido. Quando falam analisam o critério de *"uso transformativo"*, o processo envolvido está no treinamento que resulta no modelo de IAG capaz de gerar obras. Quando falam no critério de *"concorrência com a exploração econômica pelo autor"*, o processo diz respeito à etapa de inferência, pois a concorrência é exercida por obras geradas pelo sistema de IAG baseado no modelo.

⁵⁰ UNITED STATES DISTRICT COURT NORTHERN DISTRICT OF CALIFORNIA. Case 3:24-cv-05417-WHA, Document 231, Filed 06/23/25 (*Order on Fair Use*). Disponível em: <https://admin.bakerlaw.com/wp-content/uploads/2025/07/ECF-231-Order-on-Fair-Use.pdf>. Acesso em 26 out. 2025.

Outra forma de uso envolvendo inclusão de obra em base de dados foi abordada no caso *Vanderhye v. iParadigm* (2009),⁵¹ em que a Corte de Apelações norte-americana não reconheceu violação autoral no desenvolvimento do software Turnitin para detecção de plágio. Os membros da Corte entenderam ser desnecessária a autorização pelos autores para inclusão de suas obras em bases de dados com a finalidade de compará-las com novos documentos, pois o processo computacional de verificação “*não guardava qualquer relação com seu conteúdo expressivo*”. Muito embora o argumento tenha se enquadrado na doutrina do *fair use*, como uso transformativo, o aspecto distintivo aqui é a ausência de expressão da obra ou de seu sentido na comunicação autor-público. Trata-se de propósito completamente distinto, de puro processamento computacional para trazer uma informação *sobre a obra*.

Uma série de casos semelhantes de indexação de documentos na web ou indexação de conteúdo de livros encontraram desfecho semelhante, onde embora haja inclusão em base de dados, o objetivo não é disponibilizar acesso ou reproduzir a obra instaurando a relação autor-público, como por exemplo o caso *Google v. Authors Guild* (2016),⁵² no qual a Corte de Apelações entendeu haver um processo tecnológico em que a inclusão em base de dados servia para dar informações sobre a obra e não reproduzir a obra em competição com a exploração pelo autor original.

A ideia de presença ou não de “*conteúdo expressivo*” mencionada pela Corte de Apelações consiste exatamente ao conteúdo de sentido para o humano, que enfatizamos aqui ser essencial para que se fale em exploração da obra em sua individualidade. Como mostra GRIMMELMAN,⁵³ na história de desenvolvimento desse conceito e na oscilação da jurisprudência sempre esteve presente a noção de um uso da obra apenas para “*os olhos do robô*” ou se o uso envolveria a reprodução da obra para “*os olhos do humano*”. Assim, casos como o uso de papel perfurado para reprodução de música por pianola ou uso do código objeto, em oposição ao código fonte do software, inicialmente foram compreendidos como “*usos para os olhos do robô*”, mas que evoluíram posteriormente para a compreensão de que se destinavam a uma reprodução para o “*olho humano*”. O traço crucial está em saber se o processamento dos dados resultará ou não na expressividade do sentido da obra individualizada para a inteligibilidade do humano, ou seja, se estamos ou não diante de um processo que resultará na comunicação autor-público.

SAG, por sua vez, usa a metáfora bem-apanhada de um “*moinho de dados*” para essas formas de uso do dado digital, da qual a obra é uma projeção, mas que não envolve um “uso

⁵¹ A.V. ex rel. *Vanderhye v. iParadigms, LLC*, 562 F.3d 630 (4th Cir. 2009)

⁵² *Authors Guild v. Google, Inc.*, 804 F.3d 202 (2d Cir. 2015)

⁵³ GRIMMELMAN, James. Copyright for Literate Robots (May 15, 2015). 101 Iowa Law Review 657 U of Maryland Legal Studies Research Paper No. 2015-16, 2016, available at SSRN: <https://ssrn.com/abstract=2606731>

expressivo” ou “uso para consumo”.⁵⁴ O processamento computacional dos dados digitais correspondente às obras para fim diverso relação de comunicação autor-público é como colocar dados em um moinho, inteiramente sintático e sem significado para a comunicação humana da expressão individual da obra correspondente:

“Desde que a digitalização se limite a aplicações de processamento de dados que não resultem em usos expressivos ou consumptivos, infratores de obras individuais, não se coloca questão sobre obras órfãs, pois os direitos exclusivos do titular de direitos autorais se restringem aos elementos expressivos de suas obras e aos usos expressivos dessas obras.”⁵⁵

A metáfora de SAG ajusta-se bem ao uso de dados digitais resultante do *embedding* de obras protegidas para treinamento de modelos de IAG, uma vez que esse tal processamento não envolve o uso expressivo de obras individualizadas.

Para entender esse processamento, remete-se o leitor ao Anexo Técnico, que traz descrição do treinamento de modelos de IAG, explicando as razões pelas quais o modelo resultante, em si, não guarda aspectos da obra individualizada. Em síntese, tem-se que não há cópia ou armazenagem da obra, em função:

- (i) *da sistematização dos dados usados no treinamento (Seção 3)*: os dados digitalizados correspondentes a obras são já classificados por uma série de atributos ou fatores, dos quais as obras individualizadas são instâncias específicas de cada classe, e o programa de computador estabelece relações entre esses atributos que conceituam classes de indivíduos.
- (ii) *da base estatística dos modelos de IAG (Seção 4.1)*: como um modelo de aprendizado de máquina baseia-se na *frequência dos acontecimentos*, ele “valoriza” o que ocorre com mais frequência e ignora ou dilui o que é muito específico ou excepcional. Por isso, o modelo representa adequadamente o comportamento geral de uma classe estabelecendo relações entre atributos que definem aquelas classes, bem como extrai a importância relativa de cada atributo (pesos ou parâmetros) que melhor explica o agregado de dados, mas não serve para descrever ou representar cada pessoa, obra individual ou evento específico
- (iii) *da complexidade do treinamento de redes neurais profundas (Seção 4.3)*: os modelos de IAG são baseados em redes neurais profundas, com arquitetura complexa de

⁵⁴ SAG, M. Orphan works as grist for the data mill. *Berkeley Technology Law Journal* 27, 1503–50, 2012.

⁵⁵ “So long as digitization is confined to data processing applications that do not result in infringing expressive or consumptive uses of individual works, there is no orphan works problem because the exclusive rights of the copyright owner are limited to the expressive elements of their works and the expressive uses of their works” op. cit. (abstract)

funções interconectadas em várias camadas que processam, em cada interconexão, uma quantidade ínfima de “informação”. No treinamento, os dados são usados para que o modelo encontre o conjunto ideal de pesos ou parâmetros que minimize o erro na predição (ou maximize a acurácia em contextos de classificação), considerando todas as camadas e relações entre atributos, bem como as relações de ativação em um processo com milhões de iterações, o que torna praticamente impossível a compreensão pelo humano, dos inúmeros atributos e parâmetros encontrados durante o treinamento.

- (iv) *da aleatoriedade do modelo de IAG (4.3)*. Por sua natureza probabilística, o modelo também é aleatório e criativo, ou seja, cada execução pelo sistema de IAG nele baseado pode gerar variações distintas do mesmo prompt, também controladas por parâmetros usados no treinamento. É essa aleatoriedade controlada que dá origem à inovação — cada amostra gerada é uma combinação única de padrões aprendidos refletida em uma nova composição numérica em bits, cuja decodificação resulta em algo inédito, e não uma cópia de algo existente ou uma composição de trechos específicos de diferentes obras individualizadas (“colcha de retalhos”). A geração por um sistema de IAG, assim, não é reprodução, mas uma *nova composição algorítmica* obtida por inferência a partir de redes neurais profundas que representam relações matemáticas entre números binários, representando características de textos, sons, formas, cores, movimento. Tal aspecto reflete a capacidade estatística da máquina de simular criatividade sem memorizar literalmente suas fontes.

No Anexo, pode-se verificar a descrição das principais metodologias para construção de modelos fundacionais de IAG nos diferentes modais (texto, áudio, imagens, vídeos), o que o modelo “aprende” é uma função de mapeamento, ou seja, atributos ou relações indexadas a pesos (parâmetros) que capturam regularidades estatísticas dos dados usados no treinamento: co-ocorrências de tokens, padrões de estilo, estruturas de frase, padrões de formas e contornos, proporções e combinações harmônicas de cores, padrões acústicos e de ritmo, entonação, timbre etc.).

Assim, do ponto de vista técnico, no treinamento de modelos de IAG, *não há uso expressivo de qualquer obra individualizada*, apenas processamento computacional sintático e sem capacidade de expressar sentido para o humano. Os dados digitais correspondentes às obras usadas no treinamento entram em uma espécie de “moinho de dados”, no qual são extraídos padrões, conceitos, estilos,⁵⁶ resultando em um modelo estatístico geral que

⁵⁶ SARTOR, Giovanni; LAGIOIA, Francesca; CONTISSA, Giuseppe. The use of copyrighted works by AI systems: art works in the data mill. 2018. Disponível em: <https://arxiv.org/abs/1810.12211>. Acesso em: 20.10.2025.

representa o conjunto dos dados digitais e não armazena, guarda ou memoriza qualquer obra individual.⁵⁷

Essa constatação de que não há qualquer gravação, cópia ou reprodução da obra em seu uso expressivo, foi o fundamento de recente decisão adotada pela Corte de Propriedade Intelectual britânica, no caso *Getty Images versus Stable Diffusion*,⁵⁸ em que uma das questões centrais era se um modelo de IA resultante de um processo computacional de treinamento, envolvendo o ajuste de seus pesos e parâmetros com base em obras, seria uma reprodução infratora de direitos autorais. O trecho a seguir traz a resposta da juíza Joanna Smith, baseada nos pareceres técnicos dos experts, prof. Hany Farid, engenheiro e cientista da computação da Universidade da Califórnia, Berkeley e prof. Thomas Box, cientista da computação da Universidade de Freiburg:

“É verdade que os pesos do modelo são alterados durante o treinamento pela exposição a Obras Protegidas por Direito Autoral; porém, ao final desse processo, o próprio Modelo não armazena nenhuma dessas Obras Protegidas. Os pesos do modelo não constituem, por si mesmos, uma cópia infratora, nem armazenam uma cópia infratora. Eles são puramente o produto dos padrões e características que o modelo aprendeu ao longo do tempo durante o processo de treinamento. (...) Diferentemente do chip de RAM no caso Sony v. Ball, que se tornou uma cópia infratora por um curto período, na sua iteração final o Stable Diffusion não armazena nem reproduz quaisquer Obras Protegidas por Direito Autoral, e jamais o fez. O fato de que o seu desenvolvimento envolveu a reprodução de Obras Protegidas (por meio do armazenamento local de imagens e em recursos de computação em nuvem, expondo os pesos do modelo a essas imagens) é irrelevante. Além disso, o argumento de que não é necessário que um objeto que constitua uma cópia infratora continue a reter uma cópia não favorece a Getty Images, porque está implícito na palavra ‘continuar’ que, em algum momento, o objeto de fato conteve uma cópia infratora. Os pesos do modelo para cada versão do Stable Diffusion, em sua iteração final, jamais contiveram ou armazenaram uma cópia infratora.”⁵⁹

Interessante, naquele caso, foi também o debate, entre os experts trazidos pelas partes

⁵⁷GUADAMUZ, Andrés. A scanner darkly: copyright liability and exceptions in artificial intelligence inputs and outputs. *Internet Policy Review*, v. 12, n. 1, 2023. Disponível em: <https://policyreview.info/articles/secure/1771>. Acesso em: 27.10.2025

⁵⁸ *Getty Images (US) Inc & Ors v Stability AI Ltd* [2025] EWHC 2863 (Ch), Case No IL-2023-000007 (High Court of Justice, Business and Property Courts of England and Wales, Intellectual Property List (ChD), Mrs Justice Joanna Smith DBE, 4 November 2025).

⁵⁹ “it is true that the model weights are altered during training by exposure to Copyright Works, by the end of that process the Model itself does not store any of those Copyright Works; the model weights are not themselves an infringing copy and they do not store an infringing copy. They are purely the product of the patterns and features which they have learnt over time during the training process. (...) Unlike the RAM chip in Sony v Ball which became an infringing copy for a short time, in its final iteration Stable Diffusion does not store or reproduce any Copyright Works and nor has it ever done so. The fact that its development involved the reproduction of Copyright Works (through storing the images locally and in cloud computing resources and then exposing the model weights to those images) is of no relevance. Furthermore, that there is no requirement that an article which is an infringing copy must continue to retain a copy does not assist Getty Images, because it is implicit in the word “continue” that at some point the article has in fact contained an infringing copy. The model weights for each version of Stable Diffusion in their final iteration have never contained or stored an infringing copy.” Op. cit., parágrafo 600.

em relação ao fenômeno de “memorização” ou “regurgitação” que pode ocorrer na fase de inferência quando o modelo é empregado em um sistema de IAG (Ver o Anexo Técnico, item 6). No caso, algumas imagens geradas pela Stable Diffusion, apresentavam a marca d’água da Getty Images, mesmo que em imagens distintas das de propriedade da autora. O pleito ali, não dizia respeito a direito autoral mas violação de marca, por uso dela em produtos de terceiros.

Foi consenso entre os *experts*, naquele caso, tratar-se de fenômeno estatístico residual, conhecido na literatura como *memorization* ou *regurgitation*, segundo o qual o modelo, em certas condições de *prompt* ou contexto, pode reproduzir trechos de obras autorais ou dados individuais que integraram o conjunto de treino. Os *experts* destacaram tratar-se de falha ligada a *overfitting*, que pode ocorrer em modelos maiores e com *prompts* longos, com o emprego de alguma técnica para conseguir gerar trechos semelhantes ou literais, o que é corroborado pela pesquisa recente.⁶⁰ Tal fenômeno, porém, além de depender da fase de inferência, ocorre não porque o modelo possua explicitamente um “banco” das obras originais,⁶¹ mas porque durante o treino determinadas sequências — tipicamente raras ou identificáveis (por exemplo com números, nomes próprios, prefixos únicos) — geraram regiões no espaço de parâmetros/ativação que habitualmente levam à reprodução dessas sequências quando o *prompt* as aproxima (“gatilho”).

Decisão recente pelo Tribunal Regional de Munique I⁶², em ação autoral movida por sociedade de gestão coletiva contra a Open AI, entendeu, diferentemente, que o treinamento de grande modelo de linguagem configura reprodução não autorizada. (“*Vervielfältigung*”). A análise baseou-se na exceção de Text and Data Mining, prevista no § 44b UrhG, concluindo que o treinamento de modelos generativos não seria uma “análise de dados” transitória, mas uma incorporação duradoura de obras, e, portanto, fora do escopo da exceção.

O tribunal aceitou o argumento da autora de que, ao serem incorporados aos parâmetros do modelo, os textos ficariam, de certa forma, “fixados de modo perceptível” (*körperlich festgelegt*), mesmo que de “modo indireto” (*mittelbar wahrnehmbar*), pois “as características expressivas da obra podem ser novamente perceptíveis mediante os meios técnicos adequados”.⁶³ Assim, embora o Tribunal reconheça que não há cópia ou

⁶⁰ KIYOMARU, Hirokazu; SUGIURA, Issa; KAWAHARA, Daisuke; KUHASHI, Sadao. A comprehensive analysis of memorization in large language models. In: INTERNATIONAL NATURAL LANGUAGE GENERATION CONFERENCE, 17., 2024, Tokyo. Proceedings of the 17th International Natural Language Generation Conference. [S.l.]: Association for Computational Linguistics, 2024. p. 584–596.

⁶¹ HARTMANN, V., SURI, A., BINDSCHAEDLER, V., EVANS, D., TOPLE, S., & WEST, R. (2023). *SoK: Memorization in General-Purpose Large Language Models*. arXiv preprint arXiv:2310.18362. Disponível em: <https://arxiv.org/abs/2310.18362>

⁶² ALEMANHA. Landgericht München I. Sentença de 11 nov. 2025, processo 42 O 14139/24.

⁶³ “(2) Die in den Modellen körperlich festgelegten streitgegenständlichen Liedtexte können den menschlichen Sinnen mittelbar wahrnehmbar gemacht werden. Für die Vervielfältigung ausreichend ist

armazenamento de obras no modelo, fez uma espécie de equiparação jurídica a “ato de reprodução” nos termos do § 16 UrhG e do art. 2 da InfoSoc Directive, estipulando que a possibilidade de geração de trechos similares ou idênticos a obras literárias na fase de inferência traria uma “permanência da cópia” ou que a capacidade de reproduzir integralmente certas letras poderia ser considerada uma “fixação duradoura” (*dauerhafte Vervielfältigung*). Com base nessa equiparação, estende a responsabilidade por violação autoral aos desenvolvedores do modelo, ainda que a geração de trechos similares decorra de *prompts* específicos de usuários.

A decisão do Tribunal de Munique precisa ser colocada no contexto do ordenamento europeu para que a equiparação jurídica seja compreendida. A legislação autoral europeia (*EU copyright directive*), em seu art. 29, exclui da proteção autoral “atos temporários de reprodução (...) que sejam transitórios ou incidentais e parte integral de um processo tecnológico” desde que “não tenham significado econômico independente”. Em seguida, a *Directive (EU) 2019/790 (Copyright Directive for the Digital Single Market)* estipula que cópias para prospecção de dados (*Text and Data Mining-TDM*) com finalidade meramente acadêmica enquadram-se nessa exceção e a prospecção para outros fins também seria excepcionada, desde que os dados sejam retidos pelo tempo necessário para o TDM e não tenha ocorrido objeção explícita (*opt out*) pelos autores. Essa mesma disposição é incorporada no direito alemão no § 44b da Lei Alemã de Direitos Autorais e Direitos Conexos (*Urheberrechtsgesetz – UrhG*).

Assim, o que o tribunal alemão faz, nos quadros específicos daquela legislação, é tratar a possibilidade de *prompts* específicos gerarem textos similares como uma “cópia não transitória” da obra. Como cópias transitórias para TDM são exceções ao direito autoral, o tribunal faz com que a exceção não se aplique, por equiparar tais *prompts* a uma “continuidade de retenção da obra individual”.

Na verdade, a consequência proibitiva da decisão alemã somente se opera porque o direito europeu, ao estabelecer a exceção para TDM, ou seja, que o direito autoral não será aplicado se a cópia para prospecção for transitória, acaba por impor, a *contrario sensu*, que cópias não transitórias serão reprodução, mesmo que limitadas ao processamento computacional. Ou seja, o direito acaba por admitir que simples processamento computacional, por desenvolvedores que exploram dados digitais para prospecção, seja equiparado a reprodução, mesmo que não haja uso expressivo da obra.

A decisão britânica, por sua vez, separa o armazenamento da obra *para o* processamento computacional do armazenamento da obra *no* processamento, ao afirmar que

eine mittelbare Wahrnehmbarkeit, die gegeben ist, wenn das Werk unter Einsatz technischer Hilfsmittel wahrgenommen werden kann...”

a “continuidade” do armazenamento só estaria presente se as obras fossem retidas ou armazenadas no parâmetro do modelo, o que jamais ocorre (daí a ilegitimidade em se falar da fixação duradoura), permanecendo fiel ao fundamento de que apenas o uso expressivo da obra individualizada pode acarretar a proteção autoral.

Como vimos, esse fundamento é marcado tanto na concepção anglo-saxã, uma vez que o estímulo ao labor criativo com a remuneração pela exploração da obra só cabe quando há uso expressivo, quanto na concepção continental, pois é no uso expressivo da obra, na comunicação do autor com o público que se expressa sua personalidade individual.

Já no direito brasileiro, como não há estipulação legislativa semelhante à europeia para excepcionar a prospecção computacional de dados digitalizados (TDM), quando temporária, a equiparação forjada pelo tribunal alemão não tem lugar. No nosso direito, apenas a expressão individualizada da obra, seu uso expressivo, materializado por qualquer forma, pode contar como reprodução ou utilização, pois é essa a finalidade e justificativa da própria proteção autoral.

Assim, o artigo 29 da Lei de Direitos Autorais exige a autorização para qualquer tipo de utilização “da obra”, ou seja, de sua expressão individual e, como vimos, no que tange ao inc. IX (inclusão ou armazenamento em base de dados) ou ao inc. X (qualquer outra forma de utilização da obra que venha a ser inventada), tais formas de utilização devem ao menos representar parte de algum processo que resulte na comunicação autor-público por meio do acesso inteligível à obra individualizada. O mero processamento lógico-computacional do dado digital, para fim diverso, como o treinamento de modelos de IAG, que é estranho ao uso expressivo da obra, não enseja a proteção autoral e, portanto, não exige a autorização prévia de autores.

Já a geração de conteúdo similar ou de trechos idênticos que possam estabelecer o contato autor-público, individualizando a expressão original, pode constituir uma forma de reprodução, em que o sistema de IA não passa de uma ferramenta para contrafação. Mas é importante lembrar que tal contrafação é provocada pelo usuário com emprego de técnica específica, em longos prompts, que muitas vezes já contém trechos das obras, ou seja, em situações artificiais, que fogem ao uso comum de ferramentas construídas a partir de modelos de IAG.

Falar de uma “reprodução indireta” pelo desenvolvedor do modelo de IAG, além de inadequado, por alcançar agente distinto daquele que provocaria eventual violação, é desafortunado, por colocar em risco o desenvolvimento da tecnologia e os benefícios correspondentes, ligados à ampliação da capacidade de produção intelectual humana, tema que desenvolveremos no próximo item.

3. INADEQUAÇÃO DO MODELO AUTORAL PARA REMUNERAÇÃO PELO USO DE OBRAS NO TREINAMENTO DE MODELOS DE IAG

Como visto, o texto atual do PL 2338/2023, em trâmite na Câmara dos Deputados, prevê aos desenvolvedores de modelos de IAG: (i) a obrigação de listar as obras protegidas empregadas no treinamento; (ii) a proibição de utilizá-las contra a vontade manifesta de autores (chamado “*opt-out*”); (iii) a obrigação de remuneração decorrente do uso já na fase de treinamento, com exceção ao uso para atividade científica.

Tais obrigações, que compõem o modelo de remuneração autoral, oneram o desenvolvimento de sistemas de IAG tornando praticamente irreconciliáveis os objetivos de promoção da cultura nacional e de desenvolvimento tecnológico no campo da IA. Isso porque o modelo autoral traz custos significativos para os desenvolvedores de IAG, em tal monta que pode inviabilizar sua produção no País. Cabe analisar as dificuldades incorridas com a imposição de cada uma das obrigações acima mencionadas.

3.1 Dimensão qualitativa dos custos do modelo proposto pelo PL 2338/23

3.1.1. a obrigação de listar obras protegidas

Para treinar modelos de IA com desempenho competitivo, os desenvolvedores tendem a colher vastos volumes de dados com bilhões de tokens de texto, imagens, vídeo ou áudio extraídos da internet, pela prática de varredura web (*web scraping*) e utilização de “*data dumps*” públicos ou sem curadoria. Essa escala maciça já implica custos computacionais elevados (infraestrutura de armazenamento, banda, processamento, pré-processamento). Além desses custos, impor a necessidade de identificação das obras protegidas por direito autoral nesse *corpora* (textos literários, ensaios, artigos, fotografias, ilustrações, música, vídeos etc.) traria elevada complexidade, principalmente para novos entrantes ou empresas de pequeno porte.

Identificar, dentro desses volumes, quais itens são obra autoral, ou seja, têm titular protegido e não estão em domínio público ou sob licença permissiva requereria esforço técnico, operacional e jurídico para mitigar risco de infração à legislação de IA proposta.

A dificuldade reside no fato de que metadados (quem é o autor, se há licença, se é domínio público) associados aos dados digitais correspondentes às obras são frequentemente inexistentes ou são de má qualidade. Assim, é necessário investir e empregar algoritmos de classificação/filtragem que possam detectar ou inferir se determinado item é obra protegida, ou está em domínio público/licenciado, ou está explicitamente bloqueado (via *robot.txt* ou

outro mecanismo). Porém, tais algoritmos podem incorrer em falsos negativos, expondo o desenvolvedor a litigância. Assim, são necessárias iterações de curadoria, por meio de máquina ou revisão humana para sanar falsos positivos ou negativos, o que pode exigir a análise por juristas especialistas. Além disso, é necessária auditoria ou rastreamento para garantir que, se questionado por titulares, o desenvolvedor possa demonstrar que um determinado item foi incluído ou excluído com base em critérios claros (o que exige registro e logs, demandando mais capacidade computacional de armazenagem de dados).

Outro custo técnico-operacional está ligado à necessidade de rastrear quais porções de dados foram efetivamente utilizadas no treinamento, e em alguns casos “desaprender”, o que exigiria infraestrutura para mapear os dados de treinamento a versões específicas do modelo (pesada rastreabilidade), manter logs de quais itens foram excluídos ou isolados, técnicas de “*data poisoning reversal*” ou “*model unlearning*” se for requerida a remoção de influência dos dados.

Essa escala de dados e a incerteza trazida geraria uma série de custos financeiros e de tempo para desenvolvimento do projeto, tais como: horas-homem para curadoria, auditoria de dados, contabilidade de quais obras foram utilizadas, sistemas de rastreio e logs, infraestrutura para remoção (*take-down*) ou segregação e gestão de dados digitais sensíveis, reserva para contingências jurídicas, contratação de advogados especializados em direito autoral/tecnologia e contratação de seguro de responsabilidade (se disponível).

3.1.2. a obrigação de gerir o consentimento de obras identificadas

Além da identificação de obras autorais, é proposta no PL 2338/23 a exigência de gestão de consentimento de titulares (*opt-in* ou *opt-out*), o que pode ser obtido tecnicamente com a implementação de mecanismos de bloqueio (como *robot.txt*) para excluir certa parte dos dados da coleta ou do treinamento.

O mecanismo *robot.txt* (ou outros metadados machine-readable) pode permitir que criadores ou editores indiquem que seu conteúdo não deve ser incluído em operações de TDM. No entanto, na prática, a adoção desse mecanismo enfrenta uma série de dificuldades.

Primeiro, muitos sites não implementam de forma consistente *robot.txt* ou não especificam claramente “no index” para TDM. Segundo, os rastreadores (crawlers) de IA podem ignorar *robot.txt* ou haver dúvidas quanto à sua aplicabilidade.⁶⁴ Em terceiro lugar, mesmo que o *robot.txt* indique bloqueio, a remoção ou segregação retroativa dos dados já

⁶⁴ NORTON ROSE FULBRIGHT. Infringement risk relating to training a generative AI system. Global: Norton Rose Fulbright, May 2025.

coletados exige esforço operacional (filtragem, exclusão, retreinamento ou *fine-tune* incorrendo em custo). Por fim, o mecanismo de *opt-out* exige monitoramento contínuo: novos conteúdos, novos sites, novas licenças, mudança de titularidade, o que acarreta constantes custos de manutenção e governança de dados.

Assim, embora existam “ferramentas técnicas” para facilitar a gestão de consentimento, elas não eliminam o custo substancial de gerir fluxos de entrada de dados, monitorar conformidade, reagir a solicitações de retirada ou oposição dos titulares.

Alternativamente, os desenvolvedores podem recorrer a modelos de licenciamento explícito (*opt-in*), que envolve contactar os titulares, firmar contratos de uso de obras para treinamento de IA, definir remuneração ou cláusulas de remoção e rastrear quais obras foram licenciadas ou não. Tais contratos exigem a negociação de termos, como âmbito de uso (treinamento, geração, derivativos), território, remuneração, obrigações de remoção ou recall, a gestão de portfólio de direitos e manter registros, além do controle de *compliance* na organização (documentação, treinamento, checagem, estrutura de supervisão) para verificar se o uso no desenvolvimento está de acordo com os limites contratados. Portanto, custos administrativos, contábeis e jurídicos associados à gestão desses contratos.

3.1.3. a obrigação de remunerar autores ou detentores de direitos autorais

A onerosidade da obrigação de remunerar autores ou detentores de direitos autorais é imediata. Embora a remuneração aos autores possa ser justificada como forma de incentivo à produção cultural ou por justiça distributiva, a sua execução nos moldes do direito autoral, como proposto pelo PL 2338/2023, em que a remuneração é tratada como pagamento de direitos autorais pelo uso das obras para treinamento de IAGs (o que, pelo aqui exposto, está juridicamente equivocado) é praticamente inviável, seja do ponto de vista técnico, seja do ponto de vista de segurança jurídica.

Do ponto de vista técnico, deve-se lembrar que a remuneração no modelo autoral é atomizada, ou seja, remunera-se pela reprodução da obra em sua expressão individual. Ocorre que o modelo de IAG não retem ou reproduz a obra e sequer é possível mapear qual a contribuição de determinada obra para o modelo resultante.

Do ponto de vista da segurança jurídica, a remuneração em termos de indenização autoral traz, potencialmente, Judiciário como instituição adjudicação da remuneração em caso de conflitos entre autores e desenvolvedores. Diante da proliferação do treinamento de modelos de IAG, não é difícil imaginar uma corrida de autores aos tribunais para questionar diversos modelos de IAG o que, dada a complexidade do Judiciário brasileiro, a independência dos magistrados e magistradas e as dificuldades de um tema novo a ser interpretado, pode

acarretar uma série de decisões conflitantes, seja para identificar descumprimentos, seja para estipular o quantum de remuneração.

Nesse cenário de insegurança, desenvolvedores precisarão adotar decisões conservadoras quanto a reservas de contingência, contratação de advogados, negociações de acordos com uma pluralidade de autores, além de custos de litigância.

3.2 Implicações negativas para a cultura nacional e para o desenvolvimento de IAGs no Brasil

Os três conjuntos de custos trazidos pelas obrigações propostas pelo PL 2338/2023 criam três problemas principais. Primeiro, reduzir a qualidade e potencial de desenvolvimento de IAGs. Segundo, trazer barreiras significativas à entrada e ao desenvolvimento de modelos de IA, sobretudo para *startups* e atores emergentes no cenário nacional, podendo levar a concentração de mercado, sedimentando a liderança por empresas estrangeiras. Terceiro, o risco de que obras autorais nacionais não sejam utilizadas por desenvolvedores de IA, notadamente por empresas estrangeiras que tem obtido sucesso em disponibilizar e comercializar sistemas de IAG bastante populares no País, o que pode significar a alienação da cultura nacional dessas novas tecnologias em uso pelos cidadãos brasileiros, criando indesejado colonialismo digital.

A primeira consequência negativa é corolário da elevação de custos do uso do dado. Report recente da Bruegel sobre a União Europeia faz a constatação de que *“Copyright obligations reduce the quantity of available data and, through licensing requirements, increase the price of model training data.”*⁶⁵ Ou seja, mais custos para uso não expressivo de dados digitais relativos a obras naturalmente reduz a quantidade de dados disponível e o ônus financeiro para o desenvolvimento.

Conforme o modelo autoral presente no PL 2338/23, tais custos seriam incorridos, com exceção de treinamentos com a finalidade de pesquisa científica, mesmo que não haja a exploração do modelo no mercado pelo lançamento efetivo do modelo de IAG. Vale lembrar, que apenas uma fração dos modelos de IAG treinados encontram aplicações e chegam a ser lançados no mercado. Recente estudo publicado pelo MIT⁶⁶ mostra que, apesar de haver elevado investimento, apenas 40% dos projetos de IAG voltados para o público em geral são

⁶⁵ BRUEGEL. The European Union is still caught in an AI copyright bind. Bruxelas: Bruegel, 10 set. 2025. Disponível em: <https://www.bruegel.org/analysis/european-union-still-caught-ai-copyright-bind>. Acesso em: 18 nov. 2025.

⁶⁶ MASSACHUSETTS INSTITUTE OF TECHNOLOGY. The GenAI Divide: State of AI in Business 2025. Cambridge, MA: MIT NANDA, July 2025. Disponível em: https://mlq.ai/media/quarterly_decks/v0.1_State_of_AI_in_Business_2025_Report.pdf

concretizados em produtos comercializados e apenas 5% dos projetos voltados para o setor corporativo chegam a ser contratados.

Há, portanto, atualmente, elevada experimentação. Ocorre que a oneração prévia, já na fase de desenvolvimento, presente no modelo autoral proposto pelo PL 2338/23, inviabiliza essa experimentação que é a força motriz para o desenvolvimento de novas metodologias e o avanço da tecnologia. Tal força motriz seria excluída do mercado brasileiro, pela limitação imposta aos desenvolvedores que desejem realizar seus projetos no País.

Em relação a barreiras à entrada, para startups ou equipes técnicas pequenas, as três exigências representam *overhead* significativo: contratar ou treinar compliance officers, advogados, especialistas em direito autoral, implementar sistemas de registros, relatórios de auditoria, processos de retirada, “take-down” e procedimentos de monitoração (v.g., se uma obra for adicionada por um titular que não se registrou, ou se um site bloqueou uso e o dado já foi coletado).

Empresas estrangeiras já estabelecidas nesses mercados de IA ou líderes em serviços digitais que tem incorporado funcionalidades de IAG em seus produtos, já possuem escala, recursos financeiros, equipes jurídicas, infraestrutura de curadoria de dados e experiência prévia para absorver esses custos: por exemplo, já manter equipes de compliance, licenciamento, políticas de dados etc.

A alternativa competitiva para startups perante incumbentes, normalmente, liga-se agilidade, experimentação rápida, pivôs, ciclos rápidos de desenvolvimento e lançamento de versões-mínimo-viáveis (MVP). Porém, essas possíveis vantagens de agilidade em relação a líderes incumbentes são praticamente suprimidas pelas obrigações propostas pelo PL 2338/23, o que pode levar a concentração de mercado e redução da inovação no Brasil.

Por fim, é importante notar que empresas líderes estrangeiras podem desenvolver grandes modelos de linguagem e para diferentes modalidades sem necessariamente contar com obras autorais nacionais, que, mesmo assim, seriam de grande utilidade e eficiência para os usuários e consumidores brasileiros. O custo elevado, nos termos propostos pelo PL 2338/23, certamente levaria tais empresas a não usar essas obras, ou a realizar curadoria para apenas utilizar obras de domínio público, numa seleção de obras ou artistas já consagrados que possam em certa medida representar a cultura brasileira. Seria fantasioso, por outro lado, acreditar que empresas brasileiras seriam capazes de desenvolver modelos competitivos baseados em conteúdo e obras nacionais, pelos custos desproporcionalmente mais altos para alcançar qualidade semelhante ao desenvolvimento de modelos de IAG em outros países, o que praticamente pode excluir o Brasil desse mercado, a não ser, talvez, em determinados nichos para consumidores corporativos dispostos a remunerar tais custos.

Com isso, teríamos a indesejável consequência de ver a cultura nacional não incorporada nessa nova tecnologia e nova forma de consumir e produzir cultura. Considerando a crescente popularidade de assistentes de IA e sistemas geradores de conteúdo no Brasil, isso pode significar um processo de colonialismo digital, em que o acesso à tecnologia implica acesso a conteúdo treinado em dados de outra cultura, com outros valores e raízes históricas. A colonização digital vem sendo largamente debatida⁶⁷ como problema a ser enfrentado e superado por países em estágio menos avançado de desenvolvimento nessas novas tecnologias da informação, sendo a descolonização valores ético a ser perseguidos no campo digital e, em particular no campo da IA.

A proposta de remuneração presente no PL 2338/23, porém, dá passos na direção oposta, comprometendo a autonomia cultural, tecnológica e até mesmo a soberania digital no País, em completa dissonância com aquilo que preconiza o art. 219 da Constituição Federal Brasileira.

⁶⁷ MOLLEMA, Warmhold Jan Thomas. Decolonial AI as Disenclosure. *Open Journal of Social Sciences*, v. 12, n. 2, p. 574-603, 2024.

4. CONCLUSÃO

Este estudo teve por objetivo responder à seguinte questão: o uso, sem autorização, de obras protegidas na base de dados empregada para treinamento de modelos de Inteligência Artificial Generativa (IAG) configura violação autoral?

Para tanto, realizou-se ampla análise do funcionamento técnico das IAGs, conforme detalhado no Anexo Técnico, bem como do regime jurídico da Lei de Direitos Autorais e de precedentes relevantes em jurisdições de tradição continental e anglo-saxônica. A partir dessa investigação, conclui-se que o emprego de obras protegidas como dados de treinamento, sem autorização prévia, não caracteriza violação ao direito autoral.

Isso porque o objeto de proteção do direito autoral é a obra em sua expressão individualizada, materializada em suporte que possibilite comunicação entre autor e público — comunicação que pressupõe a apreensão de sentido. Tal pressuposto não se verifica no uso de dados digitalizados no treinamento de sistemas de IAG porque: (i) o processo de digitalização não envolve expressão de sentido nem comunicação autor-público; e (ii) o processamento lógico característico do treinamento produz um modelo estatístico, uma representação agregada de padrões, que não reproduz nem armazena características individuais das obras utilizadas. Nesses termos, não há uso expressivo da obra, apenas o uso dos dados digitais correspondentes, dentro de uma espécie de “moinho de dados” ininteligível para o humano, de modo que inexistente a relação autor-público necessária à incidência da proteção prevista no art. 29 da Lei de Direitos Autorais.

Sob a perspectiva do interesse coletivo, constatou-se ainda que a solução adotada pelo PL 2.338 — ao tratar o uso de obras no treinamento como matéria autoral e estabelecer mecanismos de remuneração baseados nesse modelo — não se mostra adequada para compatibilizar os objetivos de incentivo à cultura nacional e de desenvolvimento tecnológico no campo da IA.

As obrigações dele decorrentes, tais como: (i) a necessidade de identificar as obras protegidas utilizadas no treinamento; (ii) a proibição de seu uso diante de manifestação contrária dos autores (“opt-out”); e (iii) a imposição de remuneração pelo uso ainda na fase de treinamento, excetuada a pesquisa científica, acarretam custos significativos e potencialmente impeditivos à produção de sistemas de IAG no País. O resultado provável é a colisão, e não a harmonização, entre os interesses culturais e tecnológicos nacionais.

Portanto, tanto sob o prisma jurídico quanto sob a perspectiva de políticas públicas, o uso de obras protegidas para treinamento de modelos de IAG sem autorização prévia não

configura violação autoral, e o modelo proposto pelo PL 2.338 apresenta entraves relevantes que podem comprometer a própria existência e competitividade da indústria de IAG no Brasil.

5. REFERÊNCIAS

ALEMANHA. Federal Office for Information Security. Generative AI Models: Opportunities and Risks for Industry and Authorities.

ALEMANHA. Landgericht München I. Sentença de 11 nov. 2025, processo 42 O 14139/24.

ANDERSEN v. Stability AI Ltd, Case No 3:23-cv-00201, U.S. Northern District of California.

ASCENSÃO, José de Oliveira. *Direito Autoral*. 3. ed. Rio de Janeiro: Renovar, 2007

BAKER HOSTETLER. Bartz v. Anthropic. Disponível em: <https://www.bakerlaw.com/bartz-v-anthropic/>

BARAK, Aharon. *Purposive interpretation in law*. Princeton: Princeton University Press, 2011.

BARBOSA, Denis Borges. *Direito de Autor. Questões Fundamentais de Direito de Autor*. 1. ed. Lumen Juris, 2013, p. 66.

BARTZ v. Anthropic, Case No 3:24-cv-05417, Northern District of California.

BAUTISTA, Lucia. Delaware District Court Rules No Fair Use in Artificial Intelligence Case. *The Columbia Journal of Law & the Arts*. 17 ago. 2025. Disponível em: <https://journals.library.columbia.edu/index.php/lawandarts/announcement/view/816>

BENDER, E. M.; GEBRU, T.; MCMILLAN-MAJOR, A.; SHMITCHELL, S. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *FACCT* 2021.

BLOOMBERG. OpenAI to Pay Axel Springer Tens of Millions to Use News Content. Disponível em: <https://www.bloomberg.com/news/articles/2023-12-13/openai-axel-springer-ink-deal-to-use-news-content-in-chatgpt>

BRUEGEL. The European Union is still caught in an AI copyright bind. *Bruxelas: Bruegel*, 10 set. 2025. Disponível em: <https://www.bruegel.org/analysis/european-union-still-caught-ai-copyright-bind>

BROWN, Dylan. Tracking the irresistible rise of generative AI. *LexisNexis*, fev. 2024. Disponível em: <https://www.lexisnexis.co.uk/blog/future-of-law/tracking-the-irresistible-rise-of-generative-ai>

CANDIDO, Antonio. *Literatura e Sociedade*. São Paulo: Editora Nacional, 1965, pp. 44–45.

COLUMBIA ENGINEERING. Artificial Intelligence (AI) vs. Machine Learning. Disponível em: <https://ai.engineering.columbia.edu/ai-vs-machine-learning/>

COSTA NETTO, José Carlos. *Direito Autoral no Brasil*. 5. ed. São Paulo: SaraivaJur, 2025.

DODDS, Tomás; YEUNG, Wang N.; MELLADO, Claudia; LIMA-SANTOS, Mathias-Felipe. On Controlled Change: Generative AI's Impact on Professional Authority in Journalism. out. 2025.

DOI:10.48550/arXiv.2510.19792. Acesso em: 04 nov. 2025.

ESTADOS UNIDOS. New York Times v. Microsoft and OpenAI, Case No 1:23-cv-11195, U.S. District Court for the Southern District of New York.

ESTADOS UNIDOS. UMG Recordings, Inc. et al. v. Suno Inc. et al., Case No 1:24-cv-11611, U.S. District Court for the District of Massachusetts. Disponível em: <https://www.pillsburylaw.com/a/web/mASCCJ5JhVmQH2dGyXYF77/9f8pZo/suno-complaint.pdf>

ESTADOS UNIDOS. UMG Recordings Inc. et al. v. Uncharted Labs Inc., d/b/a Udio.com et al., Case No 1:24-cv-04777, U.S. District Court for the Southern District of New York. Disponível em: <https://www.pillsburylaw.com/a/web/sxgpKdHK22HVN4V7YxFcy8/9f8pZ8/udio-complaint.pdf>

FISHER, William W. "Theories of Intellectual Property." In: MUNZER, Stephen (ed.). New Essays in the Legal and Political Theory of Property. Cambridge: Cambridge University Press, 2001.

FLORIDI, Luciano. The Logic of Information: A Theory of Philosophy as Conceptual Design. Oxford: Oxford University Press, 2019.

GINSBURG, Anne. A tale of two copyrights: literary property in revolutionary France and America. Tulane Law Review, v. 64, p. 993–1031, 1990.

GOLDSTEIN, Paul. Copyright's Highway: From the Printing Press to the Cloud. Stanford University Press, 2003.

GOOGLE. Artificial Intelligence (AI) vs. Machine Learning. Disponível em: <https://cloud.google.com/learn/artificial-intelligence-vs-machine-learning>

GRAND VIEW RESEARCH. Generative AI Market (2025–2030). Disponível em: <https://www.grandviewresearch.com/industry-analysis/generative-ai-market-report>

GRIMMELMAN, James. Copyright for Literate Robots. 101 Iowa Law Review 657 (2016). Disponível em SSRN: <https://ssrn.com/abstract=2606731>.

GUADAMUZ, Andrés. A scanner darkly: copyright liability and exceptions in artificial intelligence inputs and outputs. *Internet Policy Review*, v. 12, n. 1, 2023. Disponível em: <https://policyreview.info/articles/secure/1771>

GUPTA, Shailja; RANJAN, Rajesh; SINGH, Surya Narayan. A comprehensive survey of retrieval-augmented generation (rag): Evolution, current landscape and future directions. arXiv:2410.12837, 2024.

HAGIU, Andrei; WRIGHT, Julian. Artificial intelligence and competition policy. *International Journal of Industrial Organization*, 2025.

HARTMANN, V. et al. SoK: Memorization in General-Purpose Large Language Models. arXiv:2310.18362, 2023.

HUGHES, Justin. The Philosophy of Intellectual Property. Georgetown Law Journal, 1988.

HUME, David. Tratado da natureza humana. São Paulo: Editora UNESP, 2009.

IBM. What is generative AI? Disponível em: <https://www.ibm.com/think/topics/generative-ai>

INTERNATIONAL ADVERTISING BUREAU. Generative AI Playbook for Advertising. Disponível em: https://www.iab.com/wp-content/uploads/2025/01/IAB_GenerativeAIPlaybook_January_26.pdf

JONES, R. The Myth of the Idea/Expression Dichotomy in Copyright Law. 10 Pace L. Rev. 551 (1990).

KIYOMARU, Hirokazu; SUGIURA, Issa; KAWAHARA, Daisuke; KUROHASHI, Sadao. A comprehensive analysis of memorization in large language models. INLG 2024.

LANDES, William M.; POSNER, Richard A. *The Economic Structure of Intellectual Property Law*. Cambridge: Harvard University Press, 2003.

LANDES, William M.; POSNER, Richard A. *An Economic Analysis of Copyright Law*. Journal of Legal Studies, v. 18, n. 2, p. 325–363, 1989.

LOCKE, John. Segundo tratado sobre o governo. Petrópolis: Vozes, 1994.

MARANHÃO, J. Inferências metafóricas e a reconfiguração do Direito na era digital. Revista de Direito e as Novas Tecnologias, v. 1, p. 15–30, 2018.

MARANHÃO, Juliano et al. Competition in AI Markets. Legal Wings Institute, set. 2025.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY. The GenAI Divide: State of AI in Business 2025. MIT NANDA, jul. 2025.

MCAFEE, Andrew. Generally Faster: The Economic Impact of Generative AI. abr. 2024.

McKINSEY & COMPANY. The economic potential of generative AI: The next productivity frontier. jun. 2023.

MERGES, Robert P. *Justifying Intellectual Property*. Cambridge, MA: Harvard University Press, 2011.

MOLLEMA, Warmhold Jan Thomas. Decolonial AI as Disenclosure. Open Journal of Social Sciences, v. 12, n. 2, p. 574–603, 2024.

MOONSHOT NEWS. Majority using GenAI for marketing despite concerns about technology. set. 2024.

MORAN, Lyle. 54% of in-house legal professionals support generative AI usage. Thomson Reuters Legal Dive, jun. 2023.

NORTON ROSE FULBRIGHT. Infringement risk relating to training a generative AI system. Global, maio 2025.

S&P GLOBAL. GenAI funding hits record in 2024 boosted by infrastructure interest. jan. 2025.

SAG, M. Orphan works as grist for the data mill. Berkeley Tech. Law Journal, 2012.

SAMUELS, E.; DRASSINOWER, A. A Rights-Based View of the Idea/Expression Dichotomy in Copyright Law. Canadian Journal of Law & Jurisprudence, 2003.

SANTA CATARINA. Tribunal de Justiça (TJSC). 6ª Vara Cível da Comarca de Florianópolis. Ação de inexistência de débito n. 5028302-65.2025.8.24.0023, Florianópolis, 2025.

SÃO PAULO (Estado). Tribunal de Justiça (TJSP). 3ª Vara Empresarial e de Conflitos de Arbitragem da Comarca de São Paulo. Processo n. 1107237-96.2025.8.26.0100, São Paulo, 2025.

SÃO PAULO (Estado). Tribunal de Justiça (TJSP). 18ª Vara Cível da Comarca de São Paulo. Processo n. 1119021-41.2023.8.26.0100, São Paulo, 2023.

STANFORD ENCYCLOPEDIA OF PHILOSOPHY. Intellectual Property. 2022.

SARTOR, Giovanni; LAGIOIA, Francesca; CONTISSA, Giuseppe. The use of copyrighted works by AI systems: art works in the data mill. 2018. Disponível em: <https://arxiv.org/abs/1810.12211>.

UNITED STATES DISTRICT COURT NORTHERN DISTRICT OF CALIFORNIA. Case 3:24-cv-05417-WHA, Order on Fair Use. 23 jun. 2025.

WOLFSON, Stephen. The Complex World of Style, Copyright, and Generative AI. Creative Commons.

WORLD FEDERATION OF ADVERTISERS. Majority using GenAI for marketing despite concerns about technology.

ANEXO TÉCNICO

Autores

Juliano Maranhão



Professor livre-docente da Faculdade de Direito da Universidade de São Paulo, sócio titular da Maranhão & Menezes e fundador, ao lado de professores da Poli/USP e IME/USP, da Lawgorithm Research Association. Fundador e Diretor do Legal Wings Institute.

João Pedro Lima



Mestrando em Ciência da Computação pelo IME-USP. Graduado em Ciência da Computação pela Universidade Federal do Ceará. Atua como Artificial Intelligence Engineer na empresa Artemis Technologies.

SUMÁRIO

INTRODUÇÃO	3
1. DADOS ANALÓGICOS, <i>EMBEDDING</i> , DADOS DIGITAIS E INFORMAÇÃO.....	4
2. RELAÇÕES ENTRE DADOS, CLASSES E ATRIBUTOS NO PROCESSAMENTO LÓGICO DE APRENDIZADO DE MÁQUINA.....	10
3. DADOS DIGITAIS, CLASSES E ATRIBUTOS EM MODELOS CONCEITUAIS.....	12
4. APRENDIZADO DE MÁQUINA E TREINAMENTO DE MODELOS DE IA.....	14
4.1. Modelos estatísticos	14
4.2 A representação dos dados em aprendizado supervisionado e não supervisionado....	16
4.3 Redes neurais profundas.....	18
5. IAS GENERATIVAS E MODELOS DE REPRESENTAÇÃO	24
5.1. Texto.....	25
5.2. Imagem	28
5.2.1 Geração adversarial: competição criativa entre redes neurais.....	29
5.2.2. Geração por difusão: da aleatoriedade à forma.....	30
5.3 Áudio.....	32
5.4 Vídeo	36
6. “REGURGITAÇÃO” NA FASE DE INFERÊNCIA DE SISTEMAS DE IA GENERATIVA	38
6.1 Texto.....	39
6.2 Áudio.....	39
6.3 Imagem	40
6.4 Vídeo	40
7. REFERÊNCIAS.....	42

INTRODUÇÃO

PROVENIÊNCIA DE DADOS E TREINAMENTO DE IAS GENERATIVAS

No texto jurídico tratamos o dado digital como conceito específico, distinto de informação, ou de conteúdo de sentido, que são projeções daqueles dados por meio do processamento computacional. Essa distinção é importante quando fica claro que a proteção autoral diz respeito a usos não autorizados **da obra “como obra”**, ou seja, em seu **uso expressivo**, como conteúdo semântico, em um contexto de comunicação humana entre autor e público. O uso da obra como tal, que desperta a proteção autoral, contrapõe-se ao uso dos dados digitais por meio dos quais ela se projeta, que não necessariamente desperta a proteção autoral, desde que não implique a projeção da própria obra ou parte dela como resultado do processo computacional (o que seria uma forma de reprodução). Essa distinção não é possível quando tratamos de dados analógicos, perceptíveis ao humano, em que a percepção do dado e a formação do sentido é imediata. Ele aparece com a intermediação entre o humano e o dado analógico, pela digitalização e processamento computacional.

Tendo isso em mente, na Seção 1, explicamos como se dá essa intermediação e a representação computacional do dado analógico, demonstrando que o processamento do dado digitalizado é puramente sintático, mera manipulação de símbolos, e, assim sendo, não envolve captação de sentido de cada obra individual digitalizada, tampouco a compreensão de informações sobre a obra ou sobre pessoas cujos dados analógicos foram digitalizados.

Ainda, ao investigar a etapa de processamento lógico dos dados digitalizados, em particular no treinamento de modelos conforme metodologia de aprendizado de máquina (que também é a base para desenvolvimento de IAs generativas), verificamos que se tratam de generalizações de classes de indivíduos e relações entre classes (definidas por atributos ou propriedades gerais), formando conceitos abstratos representados pelo modelo computacional, de modo que o modelo resultante do treinamento não é uma representação dos indivíduos (e.g. obras, pessoas individuais) mas uma representação de atributos, conceitos e relações abstratas, com determinados pesos (cujo parâmetro é aberto e se define pelo processo de treinamento) que delimitam a importância relativa de variáveis naquelas relações conceituais para o processo de inferência.

Como a inferência é conceitual, entre atributos de indivíduos que formam classes (não entre indivíduos que são apenas instâncias desses atributos) e como as relações guardadas são entre atributos acompanhadas dos pesos que traduzem sua importância relativa na inferência, nas IAs generativas o resultado do processamento lógico é uma combinação de bits nova, cuja decodificação traz algo inédito para a apreensão de sentido pelo humano.

Assim não se pode confundir os resultados de um modelo generativo, em quaisquer de suas modalidades como uma espécie de “colcha de retalhos” com a composição ou reprodução de trechos ou partes de diversas obras particulares.

Com isso, podemos perceber que o modelo treinado é sempre geral e não diz nada sobre os dados de entrada cuja digitalização era capaz de projetar elementos individualizados (obras ou autores/pessoas) e tudo o que diz é sobre os conceitos e relações abstratas relevantes entre as classes ou atributos.

1. DADOS ANALÓGICOS, *EMBEDDING*, DADOS DIGITAIS E INFORMAÇÃO

Dados são caracteres armazenados em determinado suporte ou mídia e podem servir de base para extração de informações. De modo abstrato, podem ser definidos como quebras de uniformidade perceptíveis (pelo humano ou por máquina). A marca de tinta azul em um papel branco é percebida pela visão humana como uma quebra de continuidade e, ao formar um símbolo em determinada linguagem ou gramática, é percebido pelo humano como sinal distintivo. Os bits, sinais elétricos, por sua vez, são percebidos pelo processamento computacional (presença/ausência do sinal). Dados são, assim, registros brutos e descontextualizados que podem ser armazenados em diferentes mídias e que, ao serem detectados isoladamente, não possuem significado ou utilidade prática (ROWLEY, 2007; FLORIDI, 2019).

Por exemplo, uma sequência de números ou símbolos isolados constitui um conjunto de dados brutos, sintáticos. Uma vez combinados de forma adequada e respeitando as regras de determinada gramática ou sistema formal, um conjunto de símbolos pode gerar significado e compor informações (linguísticas, matemáticas etc.). Na percepção humana, a detecção dos dados e a extração de significado é imediata. Quando um humano competente na linguagem identifica uma sequência de letras organizada segundo as regras de determinada gramática, imediatamente apreende o sentido das palavras e das combinações de palavras em uma sentença. Quando percebe auditivamente descontinuidades em amplitudes de ondas sonoras ou vibrações capta sons, vozes etc. o que lhe traz entendimento ou emoções. Quando está diante de obra visual, como pintura, escultura ou instalação, percebe imediatamente cores, formas, relevos, profundidade e perspectivas que permitem a compreensão e despertam sensações. Também com vídeos, imediatamente com a visualização, o humano percebe, na interface, cores, formas e movimentos, aliado à narrativa ou música trazida pela sonoridade apreendida, o que compõe significado. Em todas essas mídias, a detecção dos dados e a obtenção da informação, ou seja, do conteúdo de sentido gerado pela articulação dos dados brutos, é *imediata*.

A percepção e compreensão de textos, áudios, imagens e vídeos digitalizados, por sua vez, é intermediada por computadores que codificam, isto é, transformam caracteres, amplitudes de ondas sonoras, relações entre cores e sequências de imagem e sons, em números e bits (código binário), e posteriormente decodificam a representação numérica dos dados (em bits) por meio do hardware e de equipamentos associados como CPU, GPU, DAC, monitor, alto-falante, para gerar letras, sons, formas, cores, movimento. O quadro abaixo sintetiza a intermediação com codificação e representação binária dos dados e a decodificação por equipamentos para projetar a percepção de humana de conteúdo com significado:

Mídia	Dados binários representam...	Equipamentos que decodificam	Forma percebida	
Texto	Códigos numéricos de caracteres	CPU + GPU + monitor	Letras e palavras	<p>DO BINÁRIO AO SIGNIFICADO</p> <p>Texto: Ícone de documento, bits 01000001, seta com 'Codificação / Decodificação', ícone de monitor com letra 'A'.</p> <p>Áudio: Ícone de onda sonora, bits 01001100, 10110140..., seta com 'Codificação / Decodificação', ícone de alto-falante.</p> <p>Imagem: Ícone de pixels coloridos, bits 11111111, 00000000, 00000000, seta com 'Codificação / Decodificação', ícone de monitor com imagem.</p> <p>Vídeo: Ícone de filme, bits 01000111, 010001E1..., seta com 'Codificação / Decodificação', ícone de monitor com play.</p>
Áudio	Amplitudes da onda sonora	DAC + alto-falante	Sons e vozes	
Imagem	Cores dos pixels (RGB)	GPU + monitor	Formas e cores	
Vídeo	Sequência de imagens + som	GPU + monitor + alto-falante	Movimento com som	

Quadro 1: Formas de codificação e decodificação de dados. Elaboração própria.

Ou seja, basicamente, o computador representa o mundo em números e processa sintaticamente esses números (*i.e.* processa sinais elétricos que representam números em notação binária), o hardware aliado a equipamentos converte números em sinais físicos (luz, som, vibração) e só então o cérebro humano percebe os sinais físicos no mundo, compreendendo o significado (cognitivo ou emotivo) ao articulá-los. A informação emerge como o conteúdo semântico obtido após o processamento dos dados, ou seja, surge quando dados perceptíveis pelo humano são organizados, interpretados e inseridos em um contexto, passando a fazer sentido para algum objetivo específico (FLORIDI, 2019; MARCHIORI, 2002).

Vale detalhar o processo de codificação para cada mídia, de modo a distinguir claramente, de um lado, o *processamento sintático de dados pela máquina*, que, mesmo tratando sistemas de inteligência artificial, é desprovida de significado, entendimento ou

cognição (qualquer aspecto semântico), da *percepção e compreensão do conteúdo pelo humano*, quando o resultado do processamento computacional sintático dos dados brutos é projetado pelo hardware e equipamentos acoplados.

Fique claro que computadores não entendem palavras, não enxergam imagens nem ouvem sons, apenas computam números em representação binária. Os números matemáticos são a forma mais direta de dado, que é convertido da base decimal para a o sistema binário.

Texto. O texto em uma linguagem não tem significado matemático, sendo então convertido por uma tabela de codificação (ASCII ou Unicode) a um número correspondente. Esse número em base decimal é então convertido em bits. Exemplo, a letra “A” corresponde ao número 65, no Código ASCII, que, em representação binária corresponde a 01000001. Assim, o computador não entende a letra “A” ou qualquer palavra por ela composta, apenas manipula o número 65 em representação binária. Os números representando diferentes letras são relacionados pelo programa, que faz o processamento lógico para articular sequências que formam palavras e frases. Para propiciar a leitura humana, os bits correspondentes às sequências são convertidos em sinais elétricos pela placa de vídeo para cada ponto do monitor (pixel), resultando em quebras de continuidade entre a tela e cores com formas de letras, que o cérebro humano detecta associando-as a sinais linguísticos capazes de gerar significado e, portanto, informação.

Som. O som é uma variação de pressão no ar que chega aos nossos ouvidos em forma de ondas. O computador não “ouve” nem “sente” essas vibrações — ele precisa converter o som em números que descrevam a forma da onda. O som (voz, música, ruído) é captado por um microfone que transforma vibrações em um sinal elétrico contínuo. Esse sinal elétrico é uma onda analógica contínua, mas o computador só processa números discretos. Assim é usado um Conversor Analógico Digital- ADC para converter esse sinal analógico em dados digitais, que medem a onda em intervalos regulares (amostras), onde cada amostra é convertida em um número binário (a amostra é quantizada).

Onda analógica:

~~~~~^ ^ ^ ^ ^ ^ ~~~~~^ ^ ^ ^ ^ ^ ~~~~~

Amostragem (pontos medidos):

0 0 0 0 0

Valores medidos: 0.4 0.7 0.2 -0.3 -0.6

Convertidos em binário (16 bits):

0100110010100110 0110010111011010 ...

Quanto maior o número de medidas, maior a precisão, mas também maior o tamanho do arquivo, razão pela qual essas longas sequências de números binários são comprimidas em diferentes formatos (WAV, MP3, FLAC). Na saída, a representação numérica do som é reconvertida pelo Conversor Digital-Analógico (DAC) e, por meio do alto falante e amplificador, fazem vibrar o ar, sendo que o ouvido humano percebe as variações de pressão correspondente e articula essas quebras de continuidade em voz, música ou ruído, formando assim um significado emocional ou cognitivo.

**Imagem.** O computador não “enxerga”. Uma imagem digital para um computador, é uma grade (ou matriz) de pequenos quadrados coloridos chamados pixels. Cada pixel tem uma cor representada numericamente — geralmente por três valores (R, G, B). Cada valor varia de 0 a 255 e é escrito de modo binário (8 bits por cor):

| Cor      | Vermelho<br>(R) | Verde<br>(G) | Azul<br>(B) | Binário (8 bits cada)      |
|----------|-----------------|--------------|-------------|----------------------------|
| Branco   | 255             | 255          | 255         | 11111111 11111111 11111111 |
| Preto    | 0               | 0            | 0           | 00000000 00000000 00000000 |
| Vermelho | 255             | 0            | 0           | 11111111 00000000 00000000 |

*Quadro 2: Codificação de imagens digitais. Fonte: Elaboração própria.*

O computador armazena cada pixel como três números binários (para a tonalidade de cada cor básica) e processa esses números representando relações entre diferentes quadrantes de pixels. Para se ter uma ideia da capacidade de processamento, uma imagem de 1920x1080 pixels tem mais de 2 milhões de pontos, cada um representado por três números binários. O que para os humanos é uma percepção de mudanças de cor ou de tonalidades mais claras para mais escuras ou vice-versa, no processamento computacional é traduzida por relações entre pixels, ou seja, entre os três números binários que representam cada pixel na imagem. Na decodificação, com uso da CPU e a placa de vídeo, os bits são convertidos em intensidade luminosa nos pixels e projetados na tela do monitor. A tela emite padrões de luz que permitem perceber formas cores e profundidade e a representação da imagem na mente humana emerge da interpretação desses padrões e sinais luminosos.

**Vídeo.** O vídeo nada mais é do que uma combinação entre imagens e som, no tempo. O que os humanos veem movimentos acompanhado de áudios para o computador são sequências de “frames” ou imagens (ex. 30 imagens por segundo) onde cada frame é uma imagem codificada em pixels binários. As relações entre cada um dos frames a cada segundo são, para os computadores, relações sintáticas entre números. O som é gravado separadamente e codificado. No momento da decodificação, as sequências de frames

(números e relações entre números) são sincronizadas (com os números que representam ondas sonoras) para que então esses dados computáveis projetem, por meio do hardware associado a equipamentos apropriados projetem sinais físicos perceptíveis para o humano, cuja combinação e articulação coerente forma determinado conteúdo com significado.

Em qualquer processamento computacional, portanto, temos a seguinte sequência de ações:

| Etapa                | Descrição                                                                              | Equipamento                                                        | Percepção             |
|----------------------|----------------------------------------------------------------------------------------|--------------------------------------------------------------------|-----------------------|
| Entrada de dados     | Captura de informações do ambiente físico ou do usuário.                               | Teclado, mouse, microfone, câmera, sensores.                       | Perceptível ao humano |
| Codificação binária  | Conversão dos sinais físicos em códigos digitais (bits).                               | Conversores analógico-digitais, firmware.                          | Não perceptível       |
| Processamento lógico | Manipulação e cálculo dos dados binários segundo instruções do programa.               | CPU, GPU, memória, algoritmos lógicos, simbólicos ou estatísticos. | Não perceptível       |
| Decodificação        | Conversão dos resultados binários em sinais físicos compreensíveis (luz, som, imagem). | Placas de vídeo, DACs, controladores de saída.                     | Não perceptível       |
| Saída perceptível    | Apresentação final ao usuário                                                          | Monitor, alto-falantes, impressora                                 | Perceptível ao humano |

*Quadro 3: Sequência de ações em um processamento computacional. Elaboração própria.*

Importante notar que somente na primeira e última etapas há a possibilidade de apreensão de significado, o que ocorre pela articulação mental dos dados que são perceptíveis pelo humano. As etapas de codificação, processamento lógico e decodificação, consistem em conversões de sinais físicos para digital e vice-versa e a **manipulação de símbolos** conforme determinado programa. Nelas, o processo é puramente **sintático** e não há extração de significado, compreensão ou cognição.

Assim, vamos distinguir os dados analógicos (quebras de continuidade), perceptíveis pelo humano e que permitem a apreensão imediata de conteúdo semântico, dos dados digitais (sequências de bits), perceptíveis apenas pela máquina e processados pela manipulação sintática (computação de números codificados em bits), cuja projeção, após o processo de decodificação pode gerar conteúdo semântico para o humano (informação). Essa projeção de dados digitais pela decodificação pode individualizar documentos, obras ou pessoas, ou, por outro lado, pode apenas trazer conceitos, ideias ou relações que podem ser

compreendidas por humanos, ou ainda podem trazer correlações e representações matemáticas gerais de difícil inteligibilidade para o humano.

De particular interesse para o presente estudo é a etapa de **processamento lógico**. Ela tem como elemento subjacente um programa que traz as instruções ou regras para a manipulação simbólica que a máquina irá operar. Modelos de inteligência artificial podem ser incorporados em programas de computador (*softwares*). Nas chamadas “IAs simbólicas”, as regras são pré-estabelecidas pelo programador a partir de determinada representação de um domínio de conhecimento e uma máquina de inferência (lógica formal subjacente). No aprendizado de máquina (*machine learning*), as regras são generalizadas por indução a partir de dados de entrada em um processo de aprendizado, com base em modelos estatísticos. Ao passo que nas IAs simbólicas as regras pré-estabelecidas já processam as inferências dedutivas a partir de novos inputs (de dados), nos sistemas de IA baseados em aprendizado de máquina há uma fase anterior de treinamento, onde as regras são aprendidas, e a fase posterior de inferência, em que as regras são aplicadas para novos inputs (o treinamento pode continuar também durante o emprego do sistema).

### 2. RELAÇÕES ENTRE DADOS, CLASSES E ATRIBUTOS NO PROCESSAMENTO LÓGICO DE APRENDIZADO DE MÁQUINA

O foco da nossa análise dentro do processamento lógico na construção e emprego de sistemas de IA baseados em aprendizado de máquina está especificamente na *fase de treinamento*, na qual, a partir de um modelo estatístico subjacente, é gerado um modelo de representação do conjunto de dados de entrada para determinada finalidade, por meio de construção de uma função ou funções objetivas, de maior ou menor complexidade, com parâmetros abertos (ajustados conforme os inputs recebidos) que “representam” ou “explicam” ou “extraem padrões” dos dados “observados”.

Essas funções, que comporão as regras na fase de inferência, trazem relações, não entre indivíduos, mas entre classes de indivíduos analisados a partir dos dados de entrada e, portanto, entre atributos ou propriedades que indivíduos compartilham e que definem cada classe. Ao estabelecer relações entre atributos de classes de indivíduos, as funções objetivas resultantes do treinamento são conceituais e não dizem nada sobre os indivíduos pertencentes às classes analisadas.

Assim, é importante primeiro entender a relação entre dados, classes e atributos e traremos um exemplo sobre como em um programa de IA simbólica essas relações podem ser definidas pelo programador, a partir de seu conhecimento conceitual sobre determinado domínio. (3)

Em seguida, vamos mostrar como na fase de treinamento de modelos de IA baseados em aprendizado de máquina, o modelo resultante aprende e passa a representar em sua função objetiva as relações entre classes e atributos, ou seja, conceitos gerais e abstratos, ainda que o treinamento parta de dados digitais que projetam elementos individualizados usados no treinamento (como obras ou pessoas identificadas). (4)

No passo seguinte, vale compreender o grau de complexidade de generalização presente nos modelos de redes neurais e deep learning, nos quais as IAs Generativas se baseiam, de forma a captar como o processamento computacional extrai uma multiplicidade de relações e conceitos, muitas vezes não previsíveis e mesmo inacessíveis para a cognição e para a velocidade de processamento do cérebro humano, e como os aplica nos processos de inferência. (4.3)

Vamos entender com isso, examinando os diferentes modais de IA Generativa, como as relações e combinações conceituais formadas no aprendizado, que compõe uma função matemática complexa, quando aplicadas resultam em codificações inéditas cuja projeção após o processo de decodificação, geram obras inteiramente novas, que não se confundem

## INTELIGÊNCIA ARTIFICIAL GENERATIVA: TREINAMENTO *e* DIREITO AUTORAL

com composições de trechos de obras individuais (5). Esse item será subdividido por cada modalidade: texto (5.1), áudio (5.3), imagem (5.2), vídeo (5.4).

## 3. DADOS DIGITAIS, CLASSES E ATRIBUTOS EM MODELOS CONCEITUAIS

Um modelo conceitual pode ser definido como uma representação abstrata, formal e não ambígua de um domínio de conhecimento, tratando-se de uma estrutura que descreve conceitos, suas características e os relacionamentos existentes entre eles de forma independente da implementação ou instanciação adotada (CAMPOS; CAMPOS; MEDEIROS, 2011).

Dados informáticos ou digitais não são propriamente dados brutos, pois são classificados conforme determinadas categorias conceituais. Por exemplo, podemos digitalizar os dados que individualizam os apartamentos de edifícios localizados em determinada rua de São Paulo, organizando o arquivo com os endereços específicos que individualizam cada apartamento. Ao incorporar esses dados relativos a cada apartamento como elemento, formamos uma **classe**, cuja **extensão** é dada exatamente pelos indivíduos que pertencem a esse domínio, ou seja  $\{(Rua\ X, n.3, ap.1), (Rua\ X, n.3, ap. 2), (Rua\ X, n.3, ap. n), ..., (Rua\ X, n.5, ap. 1), ..., (Rua\ X, n. m, ap. n)\}$ . Essa extensão de elementos é categorizada conceitualmente por um **atributo** ou propriedade, qual seja, *“apartamentos em edifícios na Rua X”*. Cada apartamento, que é um elemento da classe, constitui uma **instância** concreta dessa propriedade ou atributo. O atributo *“proprietários dos apartamentos dos edifícios na rua X”* forma outra classe, cuja extensão é o conjunto de instâncias daqueles que satisfazem o atributo e.g.  $\{João, Maria, Bruna, ..., José, ..., Clara\}$ . Assim, na estruturação dos dados temos **classes** de elementos que são especificadas por determinado **atributo**, dos quais cada elemento da classe é uma **instância**.

Podemos fazer relações ou associações entre instâncias desses conjuntos, por exemplo, ao definir a relação ou conjunto de pares que mapeia cada proprietário a seu respectivo apartamento  $\{(Rua\ X, n.3, ap.1, João) ...\}$ . E podemos definir novas propriedades, que especificam novos conjuntos, por exemplo *“o conjunto dos proprietários do sexo feminino”*, ou *“o conjunto dos proprietários com mais de 60 anos de idade”*, ou *“apartamentos abaixo do 5o andar”*, *“apartamentos nos edifícios até o número 50 da Rua X”*, e assim por diante. Com isso, posso encontrar relações não só entre as instâncias das respectivas classes, ou seja, entre os indivíduos (apartamentos ou pessoas) como também entre atributos, como *“pessoas acima de 60 anos são proprietários de apartamentos abaixo do 5o andar”*. E ainda, pensando em um conjunto de valores monetários associados a cada apartamento, posso fazer relações entre atributos como *“apartamentos mais altos tem valor de mercado mais elevado”*.

Os dados digitalizados e conceptualizados, cujas classes são identificadas e nominadas, conseguem ser “lidos” ou processados por máquina, para então servirem como input de programas de computador. O programa trará relações conceituais entre diferentes

atributos. Tome-se como exemplo um programa de IA simbólica elaborado para prever preços de apartamento na região de São Paulo. Na construção de um sistema de IA simbólica, cujo conhecimento e regras de inferência são pré-estabelecidos, um *expert* no domínio de aplicação, *i.e.* no mercado imobiliário da cidade, com base em sua expertise pode estabelecer relações conceituais entre fatores que influenciam o preço de apartamentos. Por exemplo, partindo de um preço base do metro quadrado por bairro, pode estruturar um modelo simples, com funções lineares, que indicam quais fatores têm relação diretamente proporcional como o preço (aumento), como o número de quartos, número de banheiros, presença de garagem, face norte, presença de área de lazer etc. e fatores que podem reduzir o valor do apartamento, como face sul, andares inferiores, ruído elevado etc.

Tais relações entre fatores relevantes e o preço de apartamento são predeterminadas, assim como sua importância relativa ou peso na predição, ou seja, cada fator tem uma influência positiva ou negativa sobre o preço final do apartamento com determinado índice, que indica qual o seu peso (ou grau de influência) para a determinação do preço de venda previsto para o apartamento. Assim, a concepção, por um programador humano, de um modelo de representação para a precificação de apartamentos em São Paulo é construída por conceitos derivados de sua expertise, que relacionam atributos gerais de apartamentos com o preço provável que o imóvel alcançaria em uma venda. A partir dessa conceptualização, no processo de inferência, o dado de entrada sobre um apartamento individual indica as suas características relevantes diante dos conceitos presentes no modelo (bairro, metragem, número de quartos, banheiros, área de lazer, face sul etc.) e o *output* entregará um preço sugerido calculado a partir da equação estipulada pelo programador, que contém os pesos relativos para cada fator. A individualização, ocorre, assim, no momento da inferência e, mesmo assim, a partir de atributos, ou seja, elementos conceituais do modelo considerados relevantes sobre o fenômeno representado.

## 4. APRENDIZADO DE MÁQUINA E TREINAMENTO DE MODELOS DE IA

A metodologia de Aprendizado de Máquina (Machine Learning) é caracterizada pela capacidade de inferir padrões e relações a partir dos dados, sem que seja necessário programar explicitamente cada regra ou instrução para resolver um problema específico. Diferentemente dos modelos de representação do conhecimento ou de IA simbólica, em que o padrão de inferência é dedutivo – partir de uma regra (*“se o apartamento é face norte, então o preço será majorado em 10%”*) e de um caso (*“o apartamento X é face norte”*), infere-se o resultado (*o apartamento X terá seu preço majorado em 10%*), os modelos de aprendizado de máquina são indutivos, ou seja, observam relações entre casos (*“ap. 1 é face norte”, “ap. 2 é face norte”, ... “ap. n é face norte”*) e resultados (*ap.1 é mais caro em 10%, ... “ap. é mais caro em 10%”*) e inferem a regra que correlaciona os dois atributos (*apartamentos face norte têm o preço elevado em 10%*).

Esse tipo de inferência indutiva ocorre no treinamento, em que o programador, em vez de fornecer regras ao programa, fornece exemplos (casos e resultados correspondentes) e utiliza algoritmos com modelos estatísticos subjacentes e parâmetros abertos, que têm a capacidade de analisar os exemplos e, a partir deles, extrair correlações entre atributos relevantes para lidar com situações semelhantes no futuro (GEITGEY, 2019); (JAMES; WITTEN; HASTIE; TIBSHIRANI, 2021). O resultado do treinamento é, assim, uma representação matemática, de natureza estatística, dos dados observados. Por esse motivo, o aprendizado de máquina emergiu historicamente como um subcampo que une a ciência da computação e a estatística, aproveitando técnicas de modelagem estatística e adaptando-as para que máquinas possam automaticamente descobrir as equações e regras que melhor explicam e preveem comportamentos a partir dos dados disponíveis (GEITGEY, 2019); (JAMES; WITTEN; HASTIE; TIBSHIRANI, 2021).

### 4.1. Modelos estatísticos

É importante entender aspecto crucial dos modelos estatísticos que formam a base de sistemas de aprendizado de máquina: tais modelos identificam apenas **padrões gerais** — tendências, médias, correlações — sem tentar explicar cada caso individual (MAYMON, 2018).

Um modelo estatístico baseado em distribuição de probabilidades é uma forma matemática de representar a realidade quando há incerteza, de modo a encontrar padrões, tendências e possibilitar predições. Ele tenta descrever como os dados de um fenômeno — como alturas de pessoas, preços de imóveis ou notas de alunos — costumam se distribuir. Imagine que observamos o peso de mil pessoas. Vamos perceber que alguns valores aparecem mais que outros — a maioria das pessoas pesa algo em torno de um certo valor

central (por exemplo, 70 kg), e poucas têm pesos muito baixos ou muito altos. Se colocarmos isso em um gráfico, a forma curva que aparece é a distribuição de probabilidades: ela mostra quais valores são mais prováveis de ocorrer.

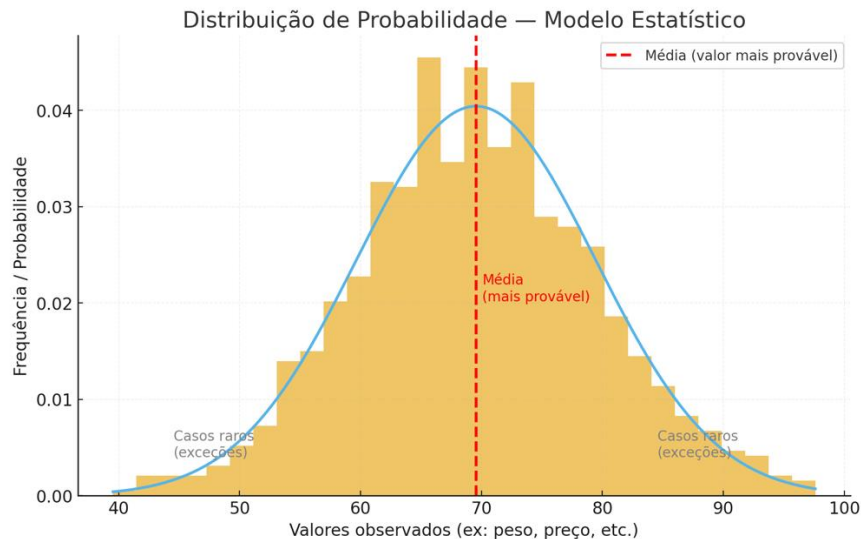


Figura 1: Gráfico de distribuição de probabilidade. Fonte: Elaboração própria.

A parte central da curva representa a *média* (ou valor típico). Os lados (as “caudas”) representam casos raros — pessoas muito mais leves ou mais pesadas. O modelo usa essa distribuição para fazer previsões ou inferências, partindo da assunção de que a maioria dos casos se comporta de modo parecido (próximo da média) e casos muito diferentes são pouco prováveis (BANDI; ADAPA; KUCHI, 2023).

Como o modelo se baseia na *frequência dos acontecimentos*, ele “valoriza” o que ocorre com mais frequência e ignora ou dilui o que é muito específico ou excepcional. Por isso, o modelo representa adequadamente o comportamento geral de um grupo mas não serve para descrever ou representar cada pessoa ou evento único. Por isso, voltando ao exemplo anterior, se aplicarmos um modelo estatístico para prever os preços de apartamentos, ele vai considerar fatores gerais como metragem, número de quartos, bairro, andar etc., mas não vai levar em conta algo singular, como a vista da janela ou o nome do proprietário, ou seus gostos na decoração. Dais especificidades individuais ficam perdidos dentro da média estatística.

#### 4.2 A representação dos dados em aprendizado supervisionado e não supervisionado

Podemos entender de modo simplificado como ocorre a construção do modelo de representação dos dados em aprendizado de máquina nas duas principais metodologias, aprendizado supervisionado e não supervisionado, usando o exemplo de predição de preços de apartamentos.

No aprendizado supervisionado, cada exemplo fornecido ao algoritmo durante o treinamento vem acompanhado de sua resposta correta, ou seja, para cada conjunto de dados de entrada existe uma variável de resposta correspondente que indica o resultado esperado.

| <i>Inputs</i>     | <i>metragem</i> | <i>quartos</i> | <i>bairro</i> | <i>banheiros</i> | <i>lazer</i> | <i>preço</i> |
|-------------------|-----------------|----------------|---------------|------------------|--------------|--------------|
| <b>1</b>          | 100m2           | 2              | Mooca         | 2                | -            | 500mil       |
| <b>2</b>          | 300m2           | 3              | Itaim         | 4                | +            | 3mi          |
| <b>...</b>        | ...             | ...            | ...           | ...              | ...          | ...          |
| <b>n</b>          | 30m2            | 1              | Jardins       | 1                | +            | 700mil       |
| <b>inferência</b> | 100m2           | 2              | Itaim         | 2                | -            | ???          |

Quadro 4 : Modelo de aprendizado supervisionado. Fonte: Elaboração própria.

Um modelo matemático simples, de equações lineares, como usado pelo programador do sistema de IA simbólica, pode ser usado como modelo subjacente para o processo de treinamento. Naquele modelo cada fator (atributo) tem uma relação direta ou indiretamente proporcional ao preço, associado a um determinado peso. A diferença, agora, é que em vez do *expert* estipular os pesos, eles ficam como parâmetros abertos.

O processo de treinamento do modelo consiste na aplicação de um algoritmo estatístico de otimização matemática, que ajusta iterativamente os pesos até minimizar ao máximo a diferença média entre o preço previsto e o preço real da amostra. Dessa forma, o modelo resultante aprende as relações entre atributos, associados aos seus pesos relativos ou grau de importância para a predição (GEITGEY, s.d.); (JAMES; WITTEN; HASTIE; TIBSHIRANI, 2021).

Portanto, o propósito primário do modelo é aprender a equação ou função matemática como modelo subjacente, que mapeia com a melhor aproximação possível, os atributos de entrada dos diferentes exemplos de apartamento aos preços de saída. Ou seja, o modelo memoriza apenas relações entre atributos associadas a pesos e não memoriza especificidades de cada apartamento, que não são relevantes estatisticamente. Para o modelo, atributos como "metragem" ou "número de quartos" são apenas números que

precisam ser considerados em certas proporções para obter a resposta (GEITGEY, 2019). No processo de inferência, o valor estimado para uma nova entrada é obtido pela multiplicação de cada atributo pelo seu respectivo peso e pela soma desses produtos (GEITGEY, s.d.; JAMES; WITTEN; HASTIE; TIBSHIRANI, 2021).

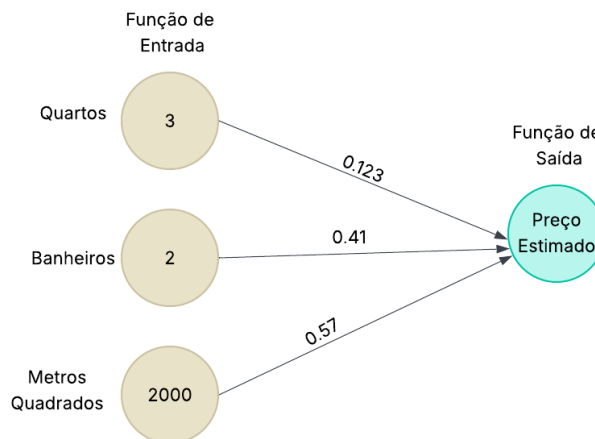


Figura 2: Exemplificação de uma rede neural ilustrativa para o problema de previsão de imóveis.

Fonte: Adaptado de (GEITGEY, 2019).

No aprendizado não supervisionado, o algoritmo recebe apenas os vetores de entrada sem qualquer informação sobre qual deveria ser a resposta ou resultado correto para aqueles dados. O objetivo do algoritmo é descobrir por conta própria a estrutura interna e as relações ocultas presentes nos dados, identificando padrões, similaridades ou agrupamentos naturais entre as observações (GEITGEY, 2019); (JAMES; WITTEN; HASTIE; TIBSHIRANI, 2021). Ele pode ser usado para encontrar atributos relevantes que ocorrem com frequência e podem ser associados em uma amostra de dados.

Para compreender como modelos não supervisionados funcionam na prática, no caso dos apartamentos, o modelo estatístico subjacente vai encontrar correlações mais frequentes entre as diferentes características dos imóveis. Imagine olharmos para as fichas técnicas de milhares de apartamentos, mas sem ver os preços – apenas tentando perceber padrões e associações naturais entre as características. O algoritmo pode descobrir, por exemplo, que o número de cômodos geralmente está correlacionado com uma maior metragem total do apartamento, formando um padrão identificável de "*apartamentos espaçosos*". Ou pode perceber que apartamentos de maior metragem ocorrem com mais frequência em determinados bairros específicos, enquanto apartamentos menores concentram-se em outras regiões, revelando assim diferentes segmentos de mercado imobiliário. Obviamente, por serem baseados em distribuições de probabilidade, tais modelos não vão guardar ou associar

agrupamentos com elementos singulares de cada imóvel, como o endereço, o nome do proprietário etc. Essa capacidade de descobrir estruturas latentes e relações ocultas nos dados de forma autônoma pode ser extremamente valiosa para explorar domínios complexos e revelar *insights* que não eram óbvios à primeira vista, mesmo para especialistas humanos na área (GEITGEY, 2019); (JAMES; WITTEN; HASTIE; TIBSHIRANI, 2021).

O que o modelo aprende, portanto, é a quantidade apropriada de "mistura" ou combinação desses números, os pesos específicos que devem ser aplicados a cada fator relevante para obter o resultado correto, seja uma previsão de preço no caso supervisionado ou uma identificação de grupos similares no caso não supervisionado. Essa característica tem uma implicação importante: *a identificação e a classificação realizadas pelo modelo baseiam-se inteiramente nas características quantificáveis que chamamos de features (características ou atributos), e não em identificadores únicos que distinguem uma observação específica de todas as outras.* Os identificadores únicos simplesmente não são necessários para a representação das correlações estatísticas e dos pesos que o modelo busca aprender, e sua inclusão não apenas seria redundante como prejudicaria a capacidade de generalização do modelo (GEITGEY, 2019); (JAMES; WITTEN; HASTIE; TIBSHIRANI, 2021).

### 4.3 Redes neurais profundas

Já pudemos observar que, em essência, os modelos de representação gerados pelo treinamento com base em aprendizado de máquina memorizam relações entre atributos (conceitos) com pesos, que são parâmetros abertos, ajustados conforme o conjunto de dados de entrada e indicam a importância relativa de cada atributo. Não memorizam aspectos específicos do dado individual (e.g. indivíduo, obra).

As redes neurais profundas também funcionam com a mesma lógica. Entendê-las, porém, permite visualizar o grau de complexidade a que podem chegar os parâmetros e relações entre atributos, o que explica porque muitas vezes os resultados do sistema trazem correlações novas e de difícil conceptualização para o humano. Daqui percebemos dois aspectos importantes: os resultados gerados não podem ser mapeados a aspectos específicos dos dados de entrada e sua explicação, em termos causais, pode ser inviável.

No exemplo simples de predição de preços de apartamentos partimos de relações lineares entre atributos. Porém, essas relações podem ser mais complexas. Por exemplo, o maior número de quartos pode ser positivo para apartamentos grandes, mas negativo para apartamentos pequenos, e o número de quartos também tem relação com o número de banheiros. Assim, podemos gerar várias estimativas otimizadas de preços com pesos distintos para cada atributo, quando se tratar de um apartamento grande ou quando se tratar de

apartamento pequeno, ou ainda, quando o apartamento tiver muitos quartos ou quando tiver poucos quartos, e assim por diante. Cada estimativa de preço terá, por sua vez, um peso distinto, também aprendido ao observar os dados de entrada, de tal modo que a estimativa de preço final é uma composição das estimas de preço conforme sua importância relativa. Além disso, algumas estimativas podem ser desconsideradas quando uma característica não for ativada (por exemplo, para um apartamento grande, não será ativada a estimativa ajustada com pesos para apartamentos pequenos). Essa ideia de ativação é similar ao funcionamento de neurônios do cérebro humano e pode ser representada matematicamente por uma composição de funções, em várias camadas. O recebimento de um dado pode alcançar determinado peso em relação a um atributo (apartamento grande) que o ativa e a informação relevante é então utilizada na camada subsequente (estimativa de preço para apartamentos grandes #1) até chegar ao resultado (composição das estimativas de preço). Na ilustração abaixo, cada atributo (metragem, quartos, banheiros) terá um peso distinto para cada estimativa de preço e cada estimativa será ativada ou não conforme os atributos presentes, influenciando o preço final, por sua vez, a partir da consideração do seu peso relativo.

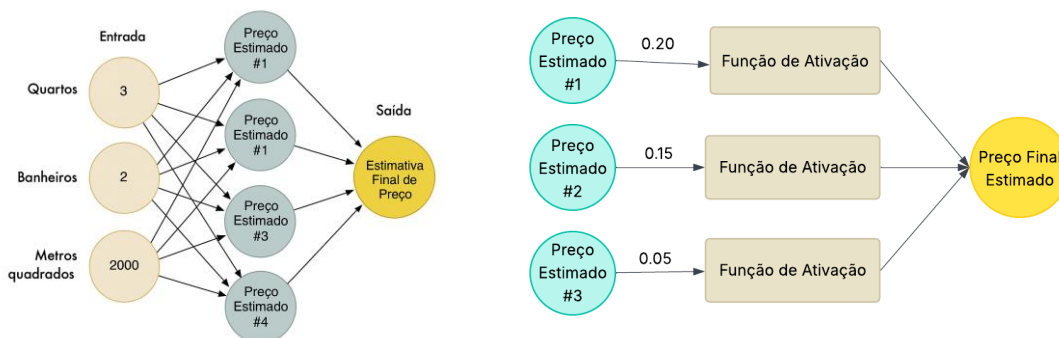


Figura 3: Exemplificação de rede neural com múltiplas camadas e atuação da função de ativação.

Fonte: Adaptado de (GEITGEY, 2019).

O modelo pode ficar cada vez mais complexo com a inserção de diversas camadas, como é comum, por exemplo, em relação a reconhecimento de imagens. O uso de redes neurais profundas, em particular Redes Neurais Convolucionais (CNNs) para reconhecimento de imagens foi um marco significativo no campo IA, oferecendo soluções para tarefas complexas anteriormente consideradas inviáveis, como o reconhecimento facial e a navegação por veículos autônomos (LI; YANG; PENG; LIU, 2020).

Como citado anteriormente, a imagem, para o sistema computacional, é inicialmente tratada como uma matriz multidimensional de valores numéricos (JAMES; WITTEN; HASTIE; TIBSHIRANI, 2021). O processamento central da CNN, a convolução, envolve a aplicação de filtros (*kernels*) que vasculham a imagem, identificando relações de coloração ou claro/escuro

em quadrantes de pixels (*patches*) para extrair características locais, como bordas ou pequenos formatos (LI; YANG; PENG; LIU, 2020); (JAMES; WITTEN; HASTIE; TIBSHIRANI, 2021). Os pesos desses filtros são aprendidos pela rede neural durante o treinamento para identificar quais padrões são mais relevantes para a tarefa de classificação (LI; YANG; PENG; LIU, 2020).

A ilustração abaixo mostra como, na classificação de uma imagem para identificar se nela há uma imagem de cachorro, diferentes atributos são delineados em diversas camadas até a classificação final:

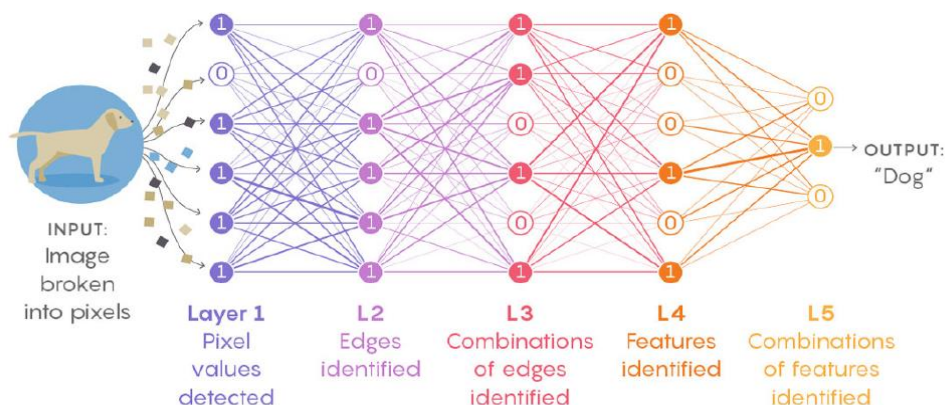


Figura 4: Exemplificação de Rede Neural Convolutacional para a tarefa de classificação de imagens.

Fonte: Maus et al. (2021).

Cada traço indica a relação de cada atributo da imagem e seu peso relativo (chamado de peso sináptico em analogia ao funcionamento do cérebro humano) para *ativar* cada “neurônio” na etapa subsequente. Os “neurônios” ativados na imagem estão com valor 1 e os não ativados com valor 0. Em cada camada, novos atributos são ativados e seu peso sináptico é relevante para os fatores que serão ativados na camada subsequente, passando de representações numéricas de pixels, para a identificação de contornos, depois a combinação de contornos que permite inferir características (orelha, focinho etc.), para a combinação de características que permite inferir se tratar de uma imagem de cachorro. Nesse processo, o sucesso computacional não vem da compreensão do significado do objeto, mas do processamento estatístico e sintático dos dados (GEITGEY, 2019). Para um sistema de *deep learning*, a imagem não representa o conceito de “cachorro”, mas uma matriz de números que codifica as intensidades de cor e luminosidade dos pixels (JAMES; WITTEN; HASTIE; TIBSHIRANI, 2021)

O objetivo do modelo de redes neurais durante o processo de treinamento é também encontrar o conjunto ideal de pesos ou parâmetros que minimize o erro na predição (ou maximize a acurácia em contextos de classificação), considerando todas as camadas e relações entre atributos, bem como as relações de ativação em um processo iterativo. Esse processo é

realizado por meio de algoritmos de otimização numérica, como o gradiente descendente (*gradient descent*), garantindo que as previsões sejam o mais precisas e consistentes possível (JAMES; WITTEN; HASTIE; TIBSHIRANI, 2021).

É importante separar a etapa de inferência, em que uma imagem individual é submetida como input para receber uma classificação pelo modelo, da etapa de treinamento, que resultará no próprio modelo, que nada mais é do que um conjunto de representações numéricas de relações entre atributos gerais associados a pesos que indicam a importância relativa de cada atributo.

Esse aspecto fica bem claro quando tomamos o processamento para reconhecimento facial. A aplicação desses modelos produz a identificação com a precisão biométrica de determinado indivíduo. Porém, o modelo não guarda qualquer aspecto individual. Na verdade, modelos de reconhecimento facial são treinados para realizar três atividades, usando redes neurais convolucionais. A primeira é identificar rostos em imagens de humanos. Para tanto, as imagens e quadrantes de pixels são convertidas em representações numéricas e o sistema é treinado para identificar o contorno do rosto, a partir de pesos atribuídos a essas relações entre representações numéricas de pixels, em diversas camadas, como na imagem de reconhecimento de um cachorro. Em seguida, como um rosto pode estar em diferentes posições, o sistema é treinado para identificar certos contornos chave de um rosto para demarcar 68 pontos cardeais, como na imagem a seguir:

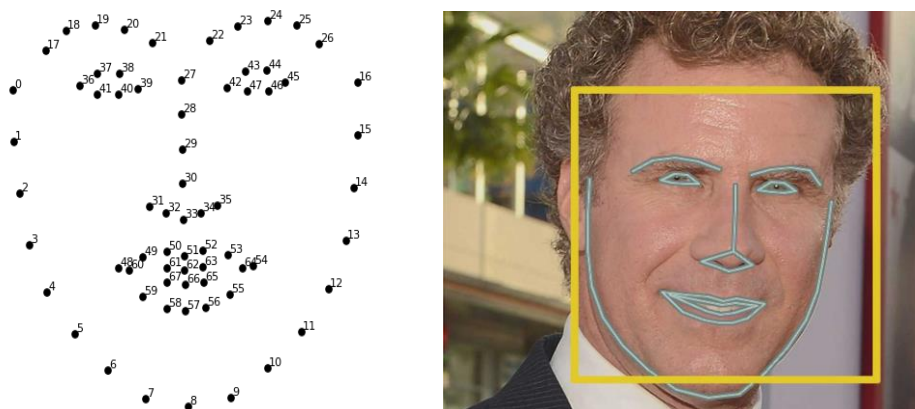


Figura 5: Processo de detecção de pontos faciais em reconhecimento facial. Fonte: (GEITGEY, 2019)

A identificação biométrica, por fim, será resultado de uma correspondência entre 128 medições entre pontos cardeais do rosto de determinada pessoa. As medições não são predeterminadas, mas aprendidas, a partir de milhares de exemplos, também com aprendizado de máquina. Assim, a imagem de um rosto, para um sistema de reconhecimento facial passa a ser um conjunto de 128 medições de distâncias entre pontos cardeais.

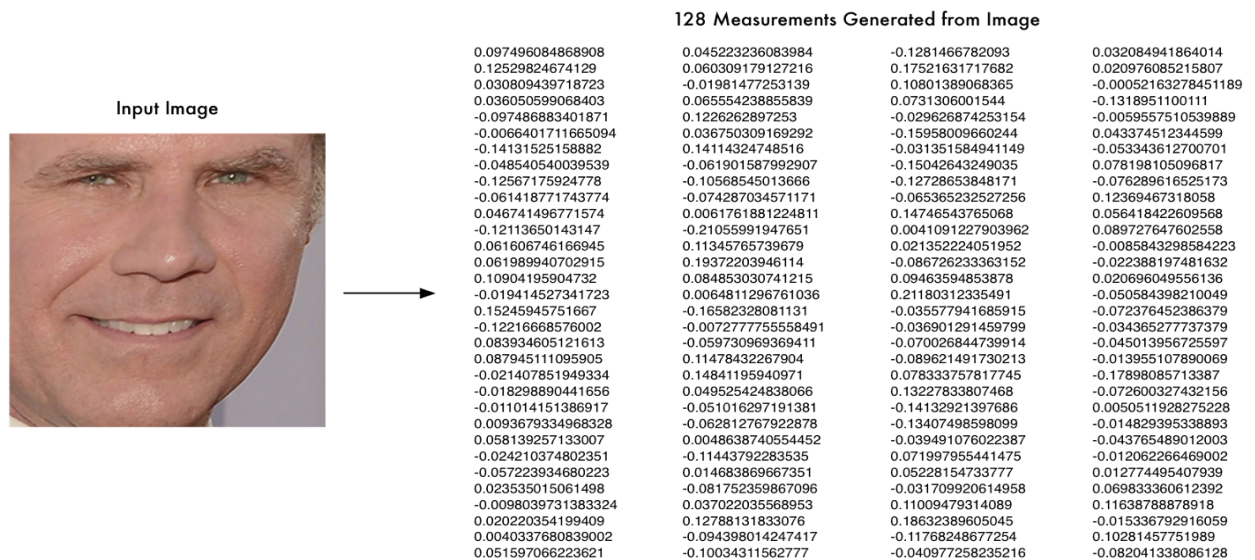


Figura 6: Exemplificação da conversão de uma imagem para vetor de características biométricas.

Fonte: (GEITGEY, 2019).

A última tarefa consiste em correlacionar as 128 medições da imagem de um rosto de uma imagem em uma base dados com o as 128 medições da imagem do rosto que se quer identificar. As medições não terão os mesmos valores, obviamente, e o modelo de rede neural, então, aprende, por milhares de exemplos, a entender quais aproximações entre a numeração que representa as 128 medições devem ser consideradas para que se obtenha um “match” (identificação) provável. Por outro lado, como a máquina, em seus cálculos, constrói relações a partir de três números em código binário associados a cada pixel, as relações aprendidas são em grande parte ininteligíveis para humanos. Ou seja, é difícil explicar por que a máquina gerou aquelas mensurações entre pontos cardeais de um rosto e não outras mensurações. Ou ainda, como chegou às aproximações entre as 128 mensurações de duas imagens que permitem considerar que se esta diante de uma identificação biométrica. O que é possível identificar apenas é o grau de acurácia, ou seja, o quanto o modelo é bem-sucedido ou não, para gerar identificações entre imagens de indivíduos.

Por fim, note-se que a identificação, com precisão biométrica, ocorre apenas na fase de inferência que compara duas imagens de determinada pessoa. Em todo o processo de treinamento, apenas atributos gerais e pesos relativos são memorizados pelo modelo: atributos do rosto humano, atributos dos contornos que permitem identificar pontos cardeais do rosto humano, mensurações diversas entre pontos cardeais, comparações entre mensurações que podem ser consideradas suficientemente próximas. Apenas esses fatores gerais são memorizados no modelo com seus respectivos pesos aprendidos no treinamento. As imagens individuais usadas no treinamento permitem que tais pesos e fatores sejam

extraídos, mas por tratarmos essencialmente de modelos estatístico, nenhum aspecto individual ou específico de cada imagem é memorizado.

## 5. IAS GENERATIVAS E MODELOS DE REPRESENTAÇÃO

Os modelos de IAGs trazem as seguintes inovações com potencial transformador, em relação aos modelos tradicionais de aprendizado de máquina:<sup>68</sup>

- (i) são sistemas genéricos ou de propósito geral, em vez de sistemas treinados para desempenhar uma tarefa específica, sendo capazes de realizar diferentes funções, a partir de treinamento com metodologias de aprendizado de máquina em imensas bases de dados;
- (ii) são capazes de gerar conteúdo novo que simula, com elevada acurácia, conteúdo de textos, imagens, sons e vídeos produzidos por humanos, sendo o conteúdo automatizado indiscernível daquele realizado por humanos. sem análise especializada;<sup>69</sup>
- (iii) possuem interfaces amigáveis, que respondem a *inputs* formulados em linguagem natural (texto ou voz) e, assim, não necessitam do conhecimento de códigos ou elementos específicos da interface de *software*.

Existem diferentes métodos e arquiteturas de modelos para que esses sistemas gerem, a partir de dados disponíveis, novos pontos de dados, capazes de projetar conteúdo inédito quando decodificados:<sup>70</sup>

- a) *Variational autoencoders* (VAEs), com uma arquitetura de codificação-decodificação e inferência variacional, em que o modelo aprende representações dos dados no *input* e gera novos dados ao selecionar amostras dos dados aprendidos;
- b) *Generative Adversarial Networks* (GANs),<sup>71</sup> compostos por um gerador que cria aleatoriamente conteúdos e um discriminador treinado para selecionar conteúdos similares a conteúdos criados por humanos;

---

<sup>68</sup> BRIGGS, J. KODNANI, D. **The potentially large effects of artificial intelligence on economic growth**, Goldman Sachs Global economic Analyst, 2023.

<sup>69</sup> COLUMBIA ENGINEERING, *Artificial Intelligence (AI) vs. Machine Learning*, Disponível em: <https://ai.engineering.columbia.edu/ai-vs-machine-learning/>. GOOGLE, *Artificial Intelligence (AI) vs. Machine Learning*, Disponível em: <https://cloud.google.com/learn/artificial-intelligence-vs-machine-learning>. IBM, *What is generative AI?*, Disponível em: <https://www.ibm.com/think/topics/generative-ai>. Acesso em 26 out. 2025

<sup>70</sup> BANDI, A.; ADAPA, P.V.S.R.; KUCHI, Y.E.V.P.K. **The Power of Generative AI: A Review of Requirements, Models, Input–Output Formats, Evaluation Metrics, and Challenges**. *Future Internet* 2023, 15, 260. <https://doi.org/10.3390/fi15080260>

<sup>71</sup> Tem-se usado o termo “modelos discriminativos” para se referir a IAGs, em oposição aos modelos descritivos tradicionais. Preferimos, porém, usar o termo modelos gerativos, tendo em vista que a principal diferença está justamente na capacidade desses modelos em gerar conteúdo novo, ao passo que o termo discriminativo se circunscreve aos modelos generativos adversariais, que compõem apenas uma das metodologias ou arquiteturas empregadas.

- c) *Diffusion Models*, que envolvem a geração de “ruídos” (*noising*) nos dados, seguida de iterações de filtragem de ruído (*denoising*) com o objetivo de aprimorar a qualidade do *output*;
- d) Modelos de fluxo normalizados aprendem distribuições probabilísticas complexas transformando iterativamente uma base de distribuição mais simples e são treinados por estimativa de maximização de proximidade;
- e) *Recurrent Neural Networks* (RNNs), geram sequências de símbolos predizendo o próximo item (token) de forma recorrente baseados em arquitetura de *codificação-decodificação*, sendo treinados por meio de aprendizagem supervisionada;
- f) *Large Language Models* (LLMs), historicamente ligados à Transformers, também são arquiteturas de codificação-decodificação que utilizam mecanismos de atenção para capturar dependências entre sequências do *input* e do *output*, de modo a refletir coerência contextual, possibilitando expressividade bastante superior e mais sofisticada em relação a RNNs.

A seguir, vamos ilustrar como operam algumas metodologias de IA Generativa, em diferentes modalidades (com exceção da geração de vídeos que é uma composição de geração de texto áudio e imagem sincronizada), de modo a perceber o ineditismo do conteúdo gerado, e sua complexidade, tendo em vista que empregam redes neurais profundas e convolucionais, o que torna praticamente impossível mapear a contribuição específica de cada obra individualizada para a construção do modelo de representação:

### 5.1. Texto

Os Chatbots de IAG utilizam-se de *Large Language Models* (LLMs) para gerar respostas em estilo de conversa aos *prompts* ou requisições submetidas pelos usuários em linguagem natural. De modo simplificado, os grandes modelos de linguagem, em geral, aprendem a prever estatisticamente a próxima melhor palavra a partir de uma sequência de palavras.<sup>72</sup> Tais modelos beneficiaram-se do desenvolvimento de redes neurais recorrentes, em sua capacidade de conectar eventos passados e futuros em uma cadeia de processamento computacional.

O grande passo para capturar e simular habilidades de escrita humanas deu-se com a estrutura de rede neural *Transformer*, em particular por sua capacidade de estabelecer

---

<sup>72</sup> GIMPEL, H.; HALL, K.; DECKER, S.; EYMANN, T.; LÄMMERMANN, L.; MÄDCHE, A.; RÖGLINGER, R.; RUINER, C.; SCHOCH, M.; SCHOOP, M.; URBACH, N.; VANDIRK, S. (2023). **Unlocking the Power of Generative AI Models and Systems such as GPT-4 and ChatGPT for Higher Education: A Guide for Students and Lecturers.** University of Hohenheim. Disponível em <[https://digital.uni-hohenheim.de/fileadmin/einrichtungen/digital/Generative AI and ChatGPT in Higher Education.pdf](https://digital.uni-hohenheim.de/fileadmin/einrichtungen/digital/Generative_AI_and_ChatGPT_in_Higher_Education.pdf)>. Acesso em 10 de jul. 2024.

*mecanismos de atenção (self-attention mechanism)* no processamento de textos.<sup>73</sup> A arquitetura dos Transformers possui um codificador (*encoder*) que, a partir de um *input* de texto, entrega como *output* não já o resultado, mas, para cada palavra ou unidade de texto, um estado oculto (*hidden state*), que será descartado no processo, e um decodificador (*decoder*) que, a partir do conjunto de estados ocultos que representam o *contexto*, produz os *outputs*. Graças a esse mecanismo é possível correlacionar estatisticamente palavras ou passagens de diferentes porções do texto de modo simultâneo ao se fazer previsões sobre textos correlatos a serem gerados no *output* (de uma tradução ou geração de texto) e não mais apenas relacionar palavras de entrada com palavras de saída de modo sequencial, como nos modelos de redes neurais recorrentes.

O mecanismo de atenção permite que o modelo avalie, para cada palavra de uma frase, quais outras palavras são mais relevantes para compreender seu sentido. Por exemplo, na frase “O gato *perseguiu* o rato porque ele era rápido”, o modelo precisa decidir a que se refere “ele” — ao gato ou ao rato? O mecanismo de atenção analisa o contexto e calcula probabilidades para cada hipótese, “pesando” as palavras anteriores para decidir qual é mais provável.

Em 2018, para aperfeiçoar seus mecanismos de busca e seu tradutor, o Google desenvolveu o modelo BERT (*Bidirectional Encoder Representation from Transformers*), com uma arquitetura de larga escala de múltiplas camadas de *transformers* bidirecionais, que revolucionou o setor de Processamento de Linguagem Natural. O modelo emprega aprendizado não supervisionado e seu pré-treinamento para “aprender” a língua ocorre por meio de dois processos, valendo-se do mecanismo de atenção:

- (i) aprender a descobrir palavras que são ocultadas aleatoriamente em textos (*Mask Language Modelling*): se o modelo recebe “O gato [mascarado] o rato”, ele testará várias possibilidades — “viu”, “comeu”, “perseguiu” — e escolherá a que tem maior probabilidade de ocorrer naquele contexto, segundo os padrões que aprendeu com milhões de textos.
- (ii) aprender se duas sentenças distintas podem ser consecutivas em um texto (*Next Sentence Prediction*): se o modelo lê “O céu está”, ele calcula, a partir de milhões de exemplos, que as próximas palavras mais prováveis são “azul”, “nublado”, “claro”, etc.

---

<sup>73</sup> Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. **Attention Is All You Need**. Available online: [https://proceedings.neurips.cc/paper\\_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf)

Após o pré-treinamento, o modelo fundacional desenvolvido pode ser refinado para tarefas mais específicas, a partir de bases de dados de diferentes domínios. As aplicações permitem a produção de palavras e construção de sentenças por predição, a partir do prompt e das palavras subsequentes geradas no processo, operação que, realizada repetidas vezes, formam mensagens e conteúdos coerentes que dão fluidez ao diálogo, até que a predição do algoritmo indique que o texto deve ser finalizado, com o conteúdo mais adequado ao conteúdo (*string* de símbolos) presente no prompt.<sup>74</sup>

Para seu desenvolvimento e produção de resultados de alto desempenho, os modelos fundacionais são treinados sobre enormes conjuntos de textos obtidos da própria internet, como páginas de sites, livros disponíveis *online* (Wikipedia e BookCorpus) e postagens em redes sociais, formando-se bases de dados com bilhões de parâmetros.<sup>75</sup>

A figura abaixo sintetiza o volume de dados, parâmetros e capacidade computacional exigida para o desenvolvimento desses modelos:

| Modelo        | Organização | Data de publicação | Dataset de treinamento <sup>1</sup> | Parâmetros                 | Poder computacional (FLOPs <sup>2</sup> ) |
|---------------|-------------|--------------------|-------------------------------------|----------------------------|-------------------------------------------|
| GPT-4         | OpenAI      | 14/03/2023         | 5.4e12<br>(5,4 trilhões)            | 1.8e12<br>(1,8 trilhão)    | 2.1e25<br>(21<br>quadrilhões)             |
| Amazon Titan  | Amazon      | 27/09/2023         | 4.0e12<br>(4 trilhões)              | 2.0e11<br>(200<br>bilhões) | 4.8e24<br>(4,8<br>quadrilhões)            |
| LLaMA 3.3 70B | MetaAI      | 05/12/2024         | 1.5e13<br>(15 trilhões)             | 7.0e10<br>(70 bilhões)     | 6.9e24<br>(6,9<br>quadrilhões)            |
| DeepSeek-V3   | DeepSeek    | 23/12/2024         | 1.5e13<br>(15 trilhões)             | 6.7e11<br>(670<br>bilhões) | 3.4e24<br>(3,4<br>quadrilhões)            |
| Pangu Ultra   | Huawei      | 09/04/2025         | 1.3e13<br>(13 trilhões)             | 1.4e11<br>(140<br>bilhões) | 1.1e25<br>(11<br>quadrilhões)             |

Figura 7: volume de dados, parâmetros e poder computacional exigidos para o desenvolvimento de alguns dos principais modelos de IA. Fonte: Epoch AI<sup>76</sup>

Com essa arquitetura complexa e bilhões de parâmetros extraídos de milhões de textos usados no treinamento, o modelo é capaz de adivinhar relações semânticas, mesmo sem compreendê-las, pois quando duas palavras aparecem com frequência nos mesmos contextos, é provável que tenham significados próximos. Assim, nos modelos de linguagem, as representações numéricas associadas a essas palavras (*embeddings*) acabam por “refletindo” relações semânticas.

<sup>74</sup> NABI, Javaid. All You Need To Know About LLM Text Generation. **Medium**, 7 ago. 2024. Disponível em: <<https://medium.com/@javaid.nabi/all-you-need-to-know-about-llm-text-generation-03b138e0ed19>>. Acesso em: 7 ago. 2024.

<sup>75</sup> SUSARLA, A.; THATCHER, R.; SARKER, S. (2023) Editorial: the janus effect of generative AI: charting the path for responsible conduct of scholarly activities in information systems. *Inf Syst Res* 34(2): 399–408. Disponível em: <<https://doi.org/10.1287/isre.2023.ed.v34.n2>> Acesso em 11 de jul. 2024. p. 400

<sup>76</sup> Data on AI Modelsort. Epoch AI, Nov. 2025. Disponível em: <https://epoch.ai/data/ai-models> Acesso em 04 de Out/2025.

Por exemplo, os *embeddings* de REI e RAINHA (usamos letras maiúsculas para indicar que na verdade temos aqui representações numéricas em código binário) terão uma diferença semelhante àquela entre HOMEM e MULHER. Da mesma forma, BRASILIA – BRASIL  $\approx$  ROMA – ITALIA expressa que a relação entre Brasília e Brasil é análoga à de Roma e Itália. Esses vetores, compostos por milhares de números, permitem que a máquina manipule representações sintáticas de significados linguísticos, por meio de representações de conceitos extraídos dos milhões de textos, de forma matemática, uma espécie de geometria do sentido.<sup>77</sup>

Como tratamos de representações conceituais matemáticas baseadas na frequência com que palavras ocorrem próximas, essa previsão é probabilística, de modo que o modelo não escolhe sempre a mesma continuação. Ele sorteia entre as opções possíveis, atribuindo maior chance às mais coerentes (conforme frequências aprendidas a partir do mecanismo de atenção). Isso explica o ineditismo do conteúdo gerado e a razão pela qual um mesmo prompt pode gerar respostas diferentes. Por exemplo: “*Escreva um poema sobre o mar*” pode gerar ora um texto romântico, ora um verso mais melancólico, conforme o modelo “explora” variações com base nas probabilidades das palavras seguintes.

Esse elemento randômico (controlado por um parâmetro chamado *temperature*) é o que torna os resultados criativos e inéditos. O modelo não repete o que leu, mas combina padrões aprendidos para formar algo novo que não ocorre nos textos usados para treinamento.

Durante o treinamento, o modelo lê bilhões de palavras provenientes de diferentes fontes. Mas ele não armazena esses textos — ele ajusta seus parâmetros internos (os chamados *pesos sinápticos* na rede neural subjacente) para representar padrões estatísticos da linguagem. Esses pesos, que podem chegar a centenas de bilhões, são números que codificam correlações: como certas palavras tendem a aparecer juntas, em que contextos, com que estrutura sintática e com representações numéricas de proximidade que refletem aquilo que para os humanos é o sentido. Assim não há “memória literal” de um texto, parágrafo ou frase específicos; há apenas traços estatísticos agregados. É impossível “abrir” o modelo e encontrar, por exemplo, o artigo de onde veio determinada expressão (ver considerações finais sobre “memorização” e “regurgitação” na aplicação de um sistema de IA Generativa).

### 5.2. Imagem

Os sistemas de geração de imagens por Inteligência Artificial baseiam-se, em sua maioria, em duas famílias de modelos generativos: as Redes Generativas Adversariais (*Generative Adversarial Networks* – GANs) e os Modelos de Difusão (*Diffusion Models*). Ambos

---

<sup>77</sup> SANTOSUOSSO, A., SARTOR, G. Decidere con l'IA: intelligenze artificiali e naturali nel diritto, Mulino, 2025.

são capazes de criar imagens novas e realistas a partir de padrões aprendidos em grandes conjuntos de dados visuais, mas operam de modos distintos.

Ambos os modelos são instâncias avançadas de redes neurais profundas, compostas por múltiplas camadas de neurônios artificiais interconectados que processam representações numéricas de dados visuais. As imagens são convertidas em matrizes numéricas – cada pixel é representado por valores que expressam sua intensidade e cor – e passam por sucessivas transformações matemáticas nas camadas da rede. O aprendizado profundo consiste justamente em ajustar os pesos sinápticos dessas conexões para que as camadas mais internas captem padrões cada vez mais complexos: bordas, formas, texturas, e finalmente estruturas como rostos, objetos e cenários. Assim, como resultado do treinamento, os números que codificavam pixels tornam-se uma representação abstrata e distribuída do conteúdo visual.

Essas representações – chamadas de *embeddings* visuais – funcionam como um espaço geométrico, análogo ao que ocorre nos modelos de linguagem com *embeddings* de palavras. Em tal espaço, imagens ou partes de imagens com características semelhantes ocupam regiões próximas, permitindo que o modelo capture relações visuais de maneira puramente numérica. Na fase de inferência, a geração de novas imagens ocorre a partir da amostragem probabilística nesse espaço latente, em que cada ponto representa uma possível combinação coerente de padrões aprendidos. Assim, o modelo não copia imagens originais, mas navega por um espaço de representações abstratas derivado das correlações numéricas extraídas de milhões de exemplos visuais, ou seja, representação do agregado de dados codificados – processo que traduz ou simula computacionalmente, em termos matemáticos, a estrutura da percepção visual humana.

O desafio técnico envolvido é duplo: primeiro, as imagens produzidas precisam apresentar um grau elevado de realismo fotográfico, assemelhando-se a fotografias autênticas; segundo o sistema precisa processar a linguagem natural na qual a imagem é requisitada e gerar imagens que correspondam ao texto, simulando uma compreensão do texto com imagem visual (SAHARIA et al., 2022).

### 5.2.1 Geração adversarial: competição criativa entre redes neurais

As GANs, introduzidas em 2014, são compostas por duas redes neurais que aprendem em oposição: (i) o *gerador*, cuja função é criar imagens sintéticas; (ii) o *discriminador*, que tenta distinguir entre imagens reais (do conjunto de treinamento) e imagens falsas (criadas pelo gerador).

Durante o treinamento, o gerador começa produzindo imagens rudimentares a partir de ruído aleatório. O discriminador aprende a identificar diferenças entre imagens “reais” e imagens “geradas”. A cada ciclo, o gerador ajusta seus parâmetros para “enganar” melhor o discriminador. Esse processo competitivo leva ambas as redes a melhorar continuamente, até que o discriminador já não consiga distinguir as imagens artificiais das reais.

O resultado é uma rede capaz de sintetizar imagens inéditas, visualmente coerentes e muitas vezes indistinguíveis de fotografias reais ou imagens geradas por humanos. Ainda que o treinamento dependa de milhões de imagens de referência, o modelo final não armazena nem reproduz nenhuma delas: os exemplos visuais servem apenas para ajustar os bilhões de parâmetros internos (os chamados *pesos sinápticos* em redes neurais) que capturam padrões estatísticos de forma, textura, cor e composição.

Assim como em modelos de linguagem, trata-se de uma aprendizagem probabilística: o gerador cria combinações de traços que têm alta probabilidade de pertencer à distribuição das imagens “reais” observadas. O que emerge é uma imagem nova, resultante de padrões e conceitos quanto a formas, contornos, cores, proporções, estilos aprendidos, e não de cópias de porções de imagens individuais. Não é possível “abrir” o modelo e identificar que pixel, objeto ou composição veio de qual imagem do treinamento — apenas se preservam correlações estatísticas de alto nível, e não registros literais.

### 5.2.2. Geração por difusão: da aleatoriedade à forma

Os modelos de difusão, hoje dominantes em aplicações como DALL·E e Stable Diffusion, seguem um princípio distinto e mais sofisticado. Eles aprendem a reverter matematicamente o processo de degradação de uma imagem. Durante o treinamento, cada imagem é progressivamente “contaminada” com ruído — até se transformar num borrão sem estrutura. A rede aprende, passo a passo, a remover o ruído e reconstruir a imagem original, compreendendo assim o caminho probabilístico entre caos e forma.

Uma vez treinado, o modelo pode gerar uma nova imagem a partir de puro ruído: começa com uma matriz de pixels aleatórios e, em centenas de interações, aplica o processo inverso aprendido — removendo ruído de forma guiada — até que surja uma imagem coerente.

O sistema parte de ruído visual aleatório e vai, passo a passo, removendo esse ruído (denoising) até formar uma imagem nítida e coerente com a descrição textual fornecida (Saharia et al. 2022).

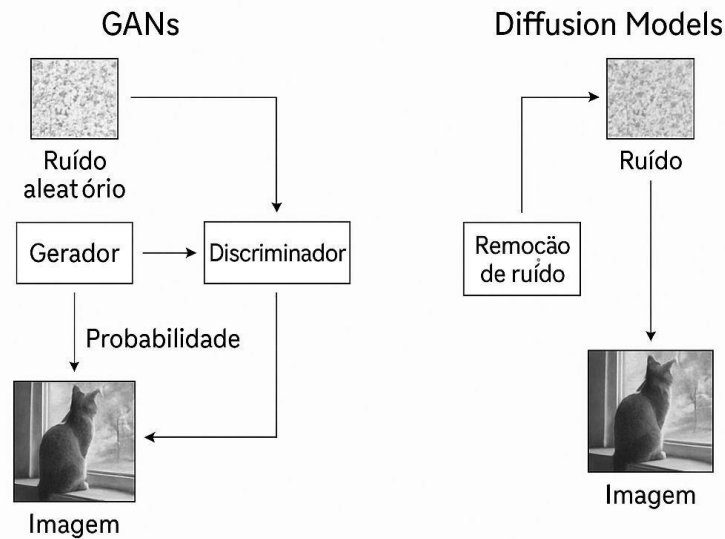


Figura 8: Comparativo entre GANs e modelos de difusão. Fonte: Elaboração própria.

Na geração de texto para imagem, se uma requisição em linguagem natural (*prompt*) for fornecida, o mecanismo de atenção cruzada (*cross-attention*) orienta o processo: as palavras do texto são associadas a padrões visuais aprendidos, como “céu azul”, “montanha”, “rosto humano”, etc. Assim, a cada passo, o modelo “pinta” a imagem de modo a maximizar a coerência entre o texto e o padrão visual que emerge. O resultado é uma imagem nova, *probabilística* e *randômica*, pois cada execução parte de um ruído inicial diferente. Ainda que se utilize o mesmo *prompt* (“um gato usando óculos escuros”), as variações do ruído inicial fazem com que surjam imagens distintas — um reflexo direto da natureza estocástica do processo de amostragem.

Uma descoberta importante no campo foi que modelos de linguagem genéricos e de larga escala – como o T5 desenvolvido pelo Google – que foram pré-treinados exclusivamente em textos (sem nunca terem sido treinados com imagens), revelam-se eficazes quando utilizados como codificadores de texto para síntese de imagens. Ainda mais interessante é que aumentar o tamanho e a capacidade do modelo de linguagem traz melhorias significativamente maiores na qualidade das imagens geradas e no alinhamento entre texto e imagem do que aumentar o tamanho do próprio modelo que gera as imagens (Saharia et al. 2022).

Durante o processo de aprendizado, os modelos de texto para imagem concentram-se em aprender a distribuição estatística dos dados de treinamento – ou seja, os padrões e características comuns presentes nas imagens utilizadas para treinar o sistema. Durante essa fase de treinamento, que exige recursos computacionais consideráveis (incluindo processadores potentes e grandes volumes de dados), o modelo aprende as relações

estatísticas e os padrões subjacentes que governam como elementos visuais se combinam para formar imagens coerentes.

Tanto nas GANs quanto nos Modelos de Difusão, o aprendizado envolve bilhões de parâmetros ajustados com base em milhões de imagens. Durante o treinamento, o modelo não “memoriza” imagens específicas, mas modifica numericamente seus parâmetros internos para representar padrões de correlação visual: como linhas e sombras formam bordas, como certas cores coexistem em determinados contextos, como determinadas proporções se repetem em rostos humanos ou objetos naturais.

Esses padrões são comprimidos em um espaço matemático multidimensional, de modo que não há mais uma correspondência entre qualquer das imagens originais usadas no treinamento e qualquer ponto específico do modelo. Por isso, é impossível mapear qual imagem de treinamento contribuiu para uma característica particular da imagem gerada. O que se conserva são relações estatísticas agregadas, e não dados literais. A geração final é fruto de amostragem probabilística nesse espaço de representações.

O modelo “escolhe” caminhos entre infinitas possibilidades, ponderando probabilidades de estrutura, cor e textura — o que explica a criatividade e o ineditismo das imagens produzidas. Assim como nos modelos de linguagem, a aleatoriedade é controlada por um parâmetro de “temperatura” (ou de “escala de ruído”), que define o quanto o modelo explorará variações menos prováveis. Temperaturas baixas produzem imagens mais previsíveis e coerentes; temperaturas altas permitem criações mais livres e experimentais. Essa aleatoriedade controlada é o que torna possível a geração de obras visuais singulares, não encontradas em nenhuma das imagens de origem — o modelo combina, transforma e generaliza padrões aprendidos para produzir algo novo e inédito.

### 5.3 Áudio

As técnicas de texto para áudio (*Text-to-Audio*) e texto para fala (*Text-to-Speech* ou *TTS*) são sistemas capazes de converter descrições textuais em fala humana natural ou gerar músicas completas com diversos estilos vocais a partir de letras ou textos descritivos. A síntese de áudio apresenta desafios únicos: a percepção auditiva humana é sensível tanto à estrutura global de uma composição musical ou discurso quanto à coerência das ondas sonoras em escalas muito estreitas de tempo. Qualquer pequena inconsistência ou artefato sonoro pode ser imediatamente percebido como artificial ou desagradável aos ouvidos humanos (HO et al., 2022); (BANDI; ADAPA; KUCHI, 2023).

Os sistemas de geração de áudio por inteligência artificial baseiam-se em redes neurais profundas capazes de aprender representações matemáticas de sons e padrões

temporais. Assim como nos modelos de linguagem ou de imagem, a IA não “memoriza” gravações específicas, mas aprende as regularidades estatísticas que estruturam o som humano – timbre, ritmo, entonação, prosódia e relações harmônicas. O treinamento ocorre sobre vastos conjuntos de gravações, que podem incluir fala, música ou sons ambientais, convertidos em representações numéricas denominadas *espectrogramas* – imagens bidimensionais que traduzem a intensidade das frequências ao longo do tempo. A partir desses espectrogramas, as redes ajustam bilhões de parâmetros internos para reconhecer correlações entre padrões acústicos e suas representações (como fonemas, notas musicais ou estilos vocais).

Nos modelos mais modernos, como os Transformers de áudio e os Modelos de Difusão de som, o processo de geração é probabilístico e randômico. Em vez de reproduzir sons existentes, o modelo estima quais amostras de áudio são mais prováveis de ocorrer após uma sequência anterior, ou, no caso dos modelos de difusão, aprende a reconstruir um som coerente a partir de ruído acústico puro. Ao gerar fala, por exemplo, o sistema parte de um texto ou *prompt* descritivo e “prediz” as ondas sonoras correspondentes, sintetizando novas vozes e inflexões que nunca existiram exatamente da mesma forma. Em música, o modelo pode criar melodias originais, combinando estilos e padrões aprendidos — um violino que segue o ritmo do jazz, ou uma voz humana com sotaque e entonação inéditos. O resultado é novo e inédito, ainda que ancorado nas distribuições estatísticas dos sons humanos e musicais usados no treinamento.

Os Transformers aplicados ao áudio seguem o mesmo princípio arquitetônico dos grandes modelos de linguagem, baseando-se no mecanismo de atenção (*attention mechanism*). Em vez de processar palavras, esses modelos lidam com fragmentos de som — pequenas janelas de amostras (*frames*) – e aprendem as relações entre elas ao longo do tempo. Durante o treinamento, o modelo aprende a prever a próxima unidade de áudio (ou o próximo bloco de espectrograma) com base no contexto anterior e posterior. A estrutura de atenção permite que ele analise dependências de longo alcance, ou seja, como um som no início de uma frase influencia a entonação no final. Isso é essencial para capturar aspectos prosódicos da fala ou continuidade harmônica na música.

Sistemas como o GANSynth utilizam também a metodologia de Redes Adversariais Generativas (GANs) e são capazes de gerar áudio de alta fidelidade, produzindo clipes musicais sintéticos. O gerador cria amostras sonoras a partir de ruído aleatório — ondas sonoras artificiais que ainda não correspondem a nenhum som real. O discriminador, por sua vez, é treinado para distinguir entre gravações reais e sons gerados. Durante o treinamento, as duas redes competem: o gerador tenta enganar o discriminador com sons cada vez mais realistas, enquanto o discriminador aprimora sua capacidade de detectar falsificações. Esse processo de

“jogo adversarial” leva o modelo a aprender padrões complexos de frequência, timbre, ritmo e textura sonora, ajustando bilhões de parâmetros internos até que os sons criados se tornem indistinguíveis dos reais.

No campo específico da síntese de fala, surgiram modelos especializados como o AdaSpeech e sistemas híbridos como o DiffGAN-TTS, que combinam diferentes arquiteturas para aumentar a qualidade da síntese. Modelos baseados em difusão como Diff-TTS, Grad-TTS e ProDiff têm focado especialmente em produzir fala cada vez mais natural e parecida com a humana, eliminando aquela qualidade “robótica” que caracterizava sistemas mais antigos e capturando as nuances sutis da fala humana natural, incluindo variações de entonação, pausas naturais e micro-variações (HO et al., 2022); (BANDI; ADAPA; KUCHI, 2023).

Os Modelos de Difusão aplicados ao som seguem o mesmo princípio probabilístico dos modelos de difusão usados em imagens, mas operam sobre sinais acústicos, isto é, variações de amplitude no tempo. O processo se baseia em dois movimentos simétricos: (i) difusão direta (*forward diffusion*) – corromper sons reais com ruído, passo a passo, até que se tornem indistinguíveis do ruído puro; (ii) difusão reversa (*reverse diffusion*) – aprender, a partir de milhões de exemplos, a remover esse ruído progressivamente e reconstruir um som coerente.

Durante o treinamento, o modelo recebe uma base de gravações de fala, música ou sons ambientais. A cada iteração, ele aplica pequenas quantidades de ruído branco sobre a forma de onda original, armazenando as diferenças entre o som puro e o som ruidoso. Esse processo é repetido milhares de vezes, em níveis graduais de degradação, até que o áudio original desapareça por completo. A rede neural, então, aprende a prever — em cada estágio — como “limpar” o ruído e recuperar o som original.

Ao final do treinamento, o modelo aprendeu a reverter matematicamente o processo de ruído, e é capaz de gerar sons novos a partir do nada, começando por uma amostra de ruído aleatório. A partir daí, em centenas de passos de *denoising*, ele transforma progressivamente o ruído em uma onda sonora coerente — uma voz, uma melodia ou um som ambiental — conforme as probabilidades aprendidas.

Quando o modelo é condicionado por texto (um *prompt* como “voz feminina lendo um poema com fundo de piano suave”), o processo é guiado por um mecanismo de atenção cruzada (*cross-attention*), que associa palavras às representações acústicas aprendidas. Assim, o sistema “pinta” o som do mesmo modo que um modelo de difusão de imagem “pinta” pixels — mas aqui o resultado é uma onda sonora. O elemento aleatório inicial (o ruído) garante que cada geração seja única: com o mesmo *prompt*, pequenas variações no ruído inicial resultam em vozes, ritmos ou timbres distintos, revelando o caráter criativo e randômico do modelo.

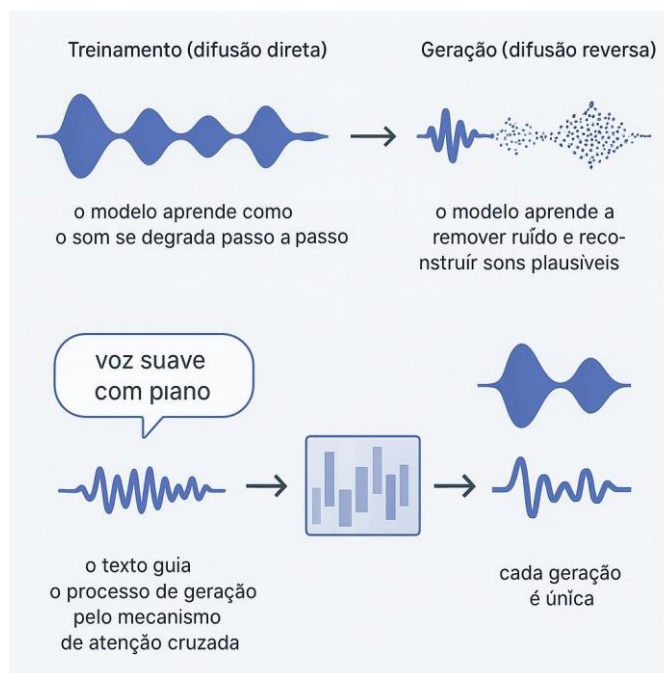


Figura 9: Síntese de áudio. Fonte: elaboração própria

O propósito fundamental desses sistemas é produzir dados sintéticos que se assemelham aos dados do mundo real e que sejam, sobretudo, realistas e diversos. Sistemas bem-sucedidos como o GANSynth exemplificam essa generalização através de sua capacidade de criar transições sonoras graduais e musicalmente coerentes entre diferentes características acústicas e manter uma identidade de timbre consistente em diferentes alturas musicais. Isso comprova que o modelo capturou a essência do estilo acústico – os princípios subjacentes que tornam um som convincente – e não apenas memorizou exemplos específicos do treinamento (HO et al., 2022).

Por sua natureza probabilística, o modelo também é aleatório e criativo: cada execução pode gerar variações distintas do mesmo prompt, controladas por parâmetros como o nível de *temperature* (grau de exploração de alternativas menos prováveis) ou a semente de ruído inicial. É essa aleatoriedade controlada que dá origem à inovação — cada amostra gerada é uma combinação única de padrões aprendidos, e não uma cópia de algo existente. A geração de áudio por IA, portanto, não é reprodução, mas composição algorítmica, em que redes neurais profundas traduzem relações matemáticas entre sons em novas expressões sonoras, refletindo a capacidade estatística da máquina de simular criatividade sem jamais possuir memória literal de suas fontes.

É, portanto, impossível mapear a contribuição de um áudio específico no resultado: as gravações usadas no treinamento servem apenas para orientar o modelo sobre as

distribuições de probabilidade do som — suas estruturas regulares, não suas ocorrências individuais.

### 5.4 Vídeo

A técnica de texto para vídeo (*text-to-video*) representa uma evolução natural dos sistemas de Inteligência Artificial Generativa para texto, áudio e imagens, permitindo que usuários criem vídeos completos a partir de simples descrições textuais. Sistemas avançados como o Imagen Video, desenvolvido pelo Google, são capazes de produzir vídeos com qualidade profissional a partir desses prompts textuais, utilizando uma arquitetura sofisticada baseada em modelos de difusão de vídeo organizados em cascata (ENGEL et al., 2019); (BANDI; ADAPA; KUCHI, 2023).

Por analogia, o funcionamento desses sistemas pode ser comparado ao trabalho de uma equipe de produção cinematográfica, onde cada membro tem uma função especializada. Primeiro, a descrição textual fornecida pelo usuário é transformada em representações matemáticas (*embeddings* de texto), utilizando um codificador de texto como o T5 – o mesmo tipo de grande modelo de linguagem que discutimos anteriormente. Esses *embeddings* funcionam como instruções codificadas que guiarão todo o processo de geração do vídeo. Em seguida, um modelo base de difusão de vídeo cria os primeiros quadros em baixa resolução e com poucos frames – seria como um esboço inicial ou um rascunho do vídeo final. A partir desse ponto, entra em ação uma sequência intercalada de modelos especializados: os modelos de Super-Resolução Espacial (SSR), que aumentam a nitidez e a qualidade visual de cada quadro, tornando-os mais detalhados e definidos; e os modelos de Super-Resolução Temporal (TSR), que preenchem os quadros intermediários entre os frames existentes, criando transições suaves e movimento fluido. Esse processo em cascata vai refinando progressivamente o vídeo até alcançar alta definição, como 1280×768 pixels rodando a 24 quadros por segundo – padrão de qualidade comparável ao que vemos em produções profissionais (ENGEL et al., 2019); (BANDI; ADAPA; KUCHI, 2023).

O processo de aprendizagem desses modelos de geração de vídeo, como o Imagen Video, concentra-se primariamente em capacitar o sistema a gerar amostras realistas e diversas, demonstrando profundo entendimento da linguagem e forte consistência temporal entre os quadros. Durante o treinamento, o foco está em fazer com que o modelo aprenda as relações estatísticas subjacentes e os padrões que governam o movimento e a evolução temporal das cenas, sendo o ideal que ele demonstre capacidade de generalizar bem para situações não vistas durante o treinamento ou com características diferentes das originais (ENGEL et al., 2019).

Um aspecto técnico particularmente revelador sobre como esses modelos aprendem relaciona-se com uma técnica chamada *classifier-free guidance*, que se mostrou crítica para a geração de amostras de alta qualidade. Quando essa técnica é aplicada, os modelos são treinados para amostrar a partir de uma distribuição inclinada para métricas de qualidade perceptual – isto é, medidas que avaliam o quanto o vídeo parece bom e realista aos olhos humanos, como o CLIP Score (que mede o alinhamento entre imagem e texto) e o CLIP R-Precision (que avalia a precisão dessa correspondência) (ENGEL et al., 2019). O ponto crucial aqui é que o objetivo não é buscar uma replicação estatística exata dos dados de treinamento, mas sim otimizar para a qualidade percebida pelo espectador humano. Essa escolha de foco revela algo fundamental sobre como o aprendizado funciona: o modelo concentra-se em capturar padrões gerais e princípios subjacentes que tornam um vídeo convincente e coerente, em oposição à memorização de instâncias específicas presentes nos dados de treinamento (ENGEL et al., 2019); (JAMES; WITTEN; HASTIE; TIBSHIRANI, 2021).

### 6. “REGURGITAÇÃO” NA FASE DE INFERÊNCIA DE SISTEMAS DE IA GENERATIVA

Como vimos na descrição das principais metodologias para construção de modelos fundacionais de IAG para de textos, áudios, imagens e vídeos, o que o modelo “aprende” é uma função de mapeamento (ou seja, atributos, conceitos ou relações indexadas a pesos, que constituem parâmetros assinalando a importância relativa dos atributos) que captura regularidades estatísticas dos dados usados no treinamento: co-ocorrências de tokens, padrões de estilo, estruturas de frase, padrões de formas e contornos, proporções e combinações harmônicas de cores, padrões acústicos e de ritmo, entonação, timbre etc.).

Então, do ponto de vista funcional, o modelo representa padrões agregados, por exemplo, *“como autores X costumam formular tal tipo de frase”, “ritmo, melodia e harmonia padrão em bossa nova”* etc. Ele não armazena *explicitamente* no seu parâmetro *“esse trecho exato da obra A da autora B”*. Em consequência, ao aplicar o modelo (ou seja, em inferência, ou geração de texto após o treino) o modelo “improvisa” com base nesses padrões: ele gera algo que “parece” aquele estilo ou aquele tipo de conteúdo, mas não aquele trecho exato ou que ele “lembre” especificamente a obra.

Apesar disso, existe a possibilidade de IAs Generativas, na fase de inferência, quando solicitadas (via *prompt*) ou inserido em determinadas condições, gerarem resultados que são muito próximos – ou mesmo idênticos – a obras autorais ou dados que estavam no seu conjunto de treino. Esse é o fenômeno chamado de “memorização” ou “regurgitação” vem sendo pesquisado e pode ocorrer ou ser provocado por requisições específicas que se direcionam a determinada obra ou conjunto de obras de um autor presente nos dados usados para treinamento. Assim, em certos casos de alta similaridade/prompt-gatilho, o modelo pode gerar resultados muito próximos ou mesmo iguais, via a dinâmica de memorização. A pesquisa tem explorado algumas hipóteses em que essa memorização ou regurgitação, sempre na fase de inferência, mediante prompts específicos, pode ocorrer.

Durante o treinamento, cada instância contribui para a atualização dos parâmetros via *back-propagation*. Exemplos que ocorrem muitas vezes, ou são “peculiares” ou de fácil “identificação” podem gerar que o modelo ajuste fortemente para “reconhecer” ou “completar” aquele exemplo.

Em modelos grandes (com muitos parâmetros) e com conjuntos de dados muito vastos, há evidência de que parte dos dados de treino (em especial os que ocorrem com alta frequência ou com identificação “única”) acabam sendo reproduzidos literalmente – ou seja, parte do modelo “sabe” que “se vier esse prompt, devo completar com esse trecho exato”.

Exemplos acadêmicos mostram que “larger models can memorize a larger portion of the data” e que “we measure exact memorization ... across model sizes”.<sup>78</sup>

No momento da aplicação (inferência), se o usuário fornecer um prompt (ou contexto) que seja muito próximo ou “ativador” de um desses trechos memorizados – por exemplo porque contém prefixo que coincide com um trecho de treino- então o modelo pode “disparar” esse trecho e reproduzi-lo (ou algo muito semelhante) *verbatim* ou quase *verbatim*.

Em cada modalidade de geração de conteúdo a memorização ou regurgitação pode ocorrer do seguinte modo:

### 6.1 Texto

Como vimos, grandes modelos de linguagem aprendem a prever o próximo token (palavra ou trecho da palavra) dado um contexto anterior, captando padrões estatísticos de sintaxe, semântica e estilo (KIYOMARU et al., 2024); (CARLINI et al., 2023).

A regurgitação, porém, pode ocorrer quando uma sequência textual rara ou distintiva (por ex. uma frase poética, um número de documento, um parágrafo jurídico) aparece com frequência nos dados, o modelo pode aprender o “caminho” completo da sequência (KIYOMARU et al., 2024); (CARLINI et al., 2023).. Se o prompt reproduz o início dessa sequência, o modelo completa com o restante literalmente, como se “lembrasse” do texto original. Por exemplo, um modelo treinado com poemas de Fernando Pessoa pode, ao receber um prompt “Ó mar salgado, quanto do teu sal...”, gerar a continuação “são lágrimas de Portugal”.

### 6.2 Áudio

Modelos de áudio (como geradores de fala ou música) mapeiam representações acústicas (espectrogramas, frequências, timbres) para gerar novos sons coerentes. O aprendizado é sobre padrões de ritmo, entonação e timbre (ZHAO et al., 2025).

A memorização pode ocorrer quando um trecho de áudio aparece repetidamente (ou é muito distinto, como a voz de uma celebridade), o modelo pode “memorizar” as características acústicas específicas. Na geração, um prompt que contenha instruções como “gerar fala semelhante a X” pode levar à reprodução parcial de fonemas, ritmo ou timbre exatos. Modelos de música podem reproduzir melodias idênticas de trechos do conjunto de treino, especialmente em músicas curtas ou populares. Um gerador de voz treinado com amostras da voz de uma pessoa pode reproduzir o mesmo timbre, pausas e entonação às

---

<sup>78</sup> Kushal Tirumala, Aram H. Markosyan, Luke Zettlemoyer Armen Aghajanyan. Memorization Without Overfitting: Analyzing the Training Dynamics of Large Language Models. 36th Conference on Neural Information Processing Systems (NeurIPS 2022).

vezes repetindo frases inteiras do *dataset*, caso receba como requisição um prompt que faça referência àquela pessoa ou a determinada melodia (ZHAO et al., 2025); (CHEN et al., 2024).

### 6.3 Imagem

Modelos como *diffusion* ou GANs aprendem a gerar imagens a partir de ruído ou relações latentes, capturando distribuições visuais gerais (formas, contornos, cores, estilos) (CHEN et al., 2024); (CARLINI et al., 2023).

A regurgitação, porém, pode ser provocada quando imagens idênticas ou muito semelhantes aparecem várias vezes no treinamento (ou são muito singulares), o modelo pode “memorizar” padrões pixel a pixel. Um prompt específico (“menina com chapéu vermelho ao pôr do sol, estilo Monet”) pode acionar a reprodução quase exata de uma foto ou pintura do dataset (CHEN et al., 2024); (CARLINI et al., 2023).

Modelos de arte generativa (*Stable Diffusion*, *Midjourney*) podem gerar quadros praticamente idênticos aos de artistas específicos usados no treinamento sem licença se forem direcionados a tal tarefa pelo prompt, que descreva o contexto de modo apropriado ou carregue imagens produzidas pelo artista cujo trecho de obra se quer gerar ou cujo estilo se quer empregar (CHEN et al., 2024); (CARLINI et al., 2023).

### 6.4 Vídeo

Modelos para geração de vídeo (como *temporal generative diffusion* ou transformers multimodais) aprendem **padrões espaço-temporais** — movimento, sequência de quadros, coerência visual ao longo do tempo (CHEN et al., 2024).

A regurgitação nesses modelos pode ser provocada na geração de vídeos dado que a base contém muitos quadros correlacionados. O vídeo, assim, amplifica a chance de que trechos curtos sejam memorizados (padrões de movimento ou cenas marcantes) (CHEN et al., 2024).

Um prompt que descreve fielmente uma cena específica (“pessoa caminhando em uma ponte de madeira com sol poente”) pode gerar a mesma sequência de quadros usada no treino (CHEN et al., 2024). Por exemplo, um modelo treinado com vídeos de esportes pode, provocado pelo prompt, como “Ronaldo Fenômeno marcando o segundo gol na final da Copa do Mundo de 2022”, pode gerar um gol idêntico ao vídeo da final da copa do mundo presente no dataset.

O fenômeno de regurgitação ou memorização, porém, não significa que o modelo de representação armazena um “índice” do dado original e o reproduz. Mas sim que o perfil dos

parâmetros e a dinâmica de decodificação favorecem a emergência desse trecho original em certas condições de prompt/contexto, o que ocorrerá somente na fase de inferência (CHEN et al., 2024).

## 7. REFERÊNCIAS

BASTI, G. Do Computers "Have Syntax, But No Semantics"? *Minds and Machines*, v. 31, p. 551-571, 2021. DOI: 10.1007/s11023-021-09564-9.

BAKER HOSTETLER. Bartz v. Anthropic. Disponível em: <https://www.bakerlaw.com/bartz-v-anthropic/>. Acesso em 20 out. 2025.

BANDI, Ajay; ADAPA, Pydi Venkata Satya Ramesh; KUCHI, Yudu Eswar Vinay Pratap Kumar. The Power of Generative AI: A Review of Requirements, Models, Input-Output Formats, Evaluation Metrics, and Challenges. *Future Internet*, v. 15, n. 8, p. 260, 2023. DOI: 10.3390/fi15080260.

BAUTISTA, Lucia. *Delaware District Court Rules No Fair Use in Artificial Intelligence Case*. The Columbia Journal of Law & the Arts. August 17, 2025. Disponível em: <https://journals.library.columbia.edu/index.php/lawandarts/announcement/view/816>. Acesso em 20 out. 2025.

BITTAR, Carlos Alberto. *Direito de Autor*. 7ª ed., Rio de Janeiro: Editora Forense, 2019.

BLOOMBERG. *OpenAI to Pay Axel Springer Tens of Millions to Use News Content*. <https://www.bloomberg.com/news/articles/2023-12-13/openai-axel-springer-ink-deal-to-use-news-content-in-chatgpt>. Acesso em 24 out. 2025.

BROWN, Dylan. *Tracking the irresistible rise of generative AI*. LexisNexis. Fevereiro 2024. Disponível em: <https://www.lexisnexis.co.uk/blog/future-of-law/tracking-the-irresistible-rise-of-generative-ai>.

BRUELL, Alexandra. *Amazon to Pay New York Times at Least \$20 Million a Year in AI Deal*. Wall Street Journal. Julho 2025. Disponível em: <https://www.wsj.com/business/media/amazon-to-pay-new-york-times-at-least-20-million-a-year-in-ai-deal-66db8503>. Acesso em 9 out. 2025.

BOYLE, James. *A Theory of Law and Information: Copyright, Spleens, Blackmail, and Insider Trading*. California Law Review, Vol. 80, No. 6, December 1992. Disponível em: [https://scholarship.law.duke.edu/cgi/viewcontent.cgi?article=1136&context=faculty\\_scholarship](https://scholarship.law.duke.edu/cgi/viewcontent.cgi?article=1136&context=faculty_scholarship). Acesso em 19 out. 2025.

CAMPOS, M. L. A.; CAMPOS, L. M.; MEDEIROS, J. S. A representação de domínios de conhecimento e uma teoria de representação: a ontologia de fundamentação. *Informação & Informação*, Londrina, v. 16, n. 3, p. 140-164, jan./jun. 2011. DOI: 10.5433/1981-8920.2011v16n3p140.

CARLINI, Nicholas; HAYES, Jamie; NASR, Milad; JAGIELSKI, Matthew; SEHWAG, Vikash; TRAMÈR, Florian; BALLE, Borja; IPPOLITO, Daphne; WALLACE, Eric. Extracting Training Data from Diffusion Models. arXiv preprint arXiv:2301.13188, jan. 2023.

CHEN, Chen; LIU, Enhuai; LIU, Daochang; SHAH, Mubarak; XU, Chang. Investigating memorization in video diffusion models. arXiv preprint, arXiv:2410.21669, 29 out. 2024.

COSTA NETTO, José Carlos. *Direito Autoral no Brasil*. 3ª ed., São Paulo: Saraiva Educação, 2019.

DUAN et al. *Uncovering Latent Memories: Assessing Data Leakage and Memorization Patterns in Frontier AI Models*. Julho 2024. <https://doi.org/10.48550/arXiv.2406.14549>.

FELIN, Teppo; HOLWEG, Matthias. *Theory Is All You Need: AI, Human Cognition, and Causal Reasoning*. Strategy Science “Theory-Based View” Conference, Bocconi University. Disponível em: [https://www.bu.edu/dbi/files/2024/08/FelinHolwegAug2024\\_SSRN.pdf](https://www.bu.edu/dbi/files/2024/08/FelinHolwegAug2024_SSRN.pdf). Acesso em 27 out. 2025.

FLORIDI, L. Semantic Conceptions of Information. In: ZALTA, E. N. (Ed.). *Stanford Encyclopedia of Philosophy*. Summer 2019 Edition. Stanford: Stanford University, 2019. Disponível em: <https://plato.stanford.edu/archives/sum2019/entries/information-semantic/>. Acesso em: 19 out. 2025.

GÜL, O. M.; UYSAL, E. A Theory of Semantic Communication. *IEEE Transactions on Information Theory*, 2024. DOI: 10.1109/TIT.2024.

HOFFMANN et al., *Training Compute-Optimal Large Language Models*, 36th Conference on Neural Information Processing Systems (NeurIPS 2022). Disponível em: [https://papers.nips.cc/paper\\_files/paper/2022/file/c1e2faff6f588870935f114ebe04a3e5-Paper-Conference.pdf](https://papers.nips.cc/paper_files/paper/2022/file/c1e2faff6f588870935f114ebe04a3e5-Paper-Conference.pdf)

KIYOMARU, Hirokazu; SUGIURA, Issa; KAWAHARA, Daisuke; KUROHASHI, Sadao. A comprehensive analysis of memorization in large language models. In: INTERNATIONAL NATURAL LANGUAGE GENERATION CONFERENCE, 17., 2024, Tokyo. *Proceedings of the 17th International Natural Language Generation Conference*. [S.l.]: Association for Computational Linguistics, 2024. p. 584–596.

MAUS, André; RAZERA, Tatiele; KUNTZ, Roney; JUNIOR, Valdir; JUNIOR, Ademir. Contagem e classificação de veículos por visão computacional. 2021. p. 49–56. DOI: 10.14210/cotb.v12.p049-056.

MARCHIORI, P. Z. A ciência e a gestão da informação: compatibilidades no espaço profissional. *Ciência da Informação*, Brasília, v. 31, n. 2, p. 72-79, maio/ago. 2002.

OLIVEIRA, V. N. P. de; ALMEIDA, M. B.; RAMALHO, F. A. Um roteiro para avaliação ontológica de modelos de sistemas de informação. *Perspectivas em Ciência da Informação*, Belo Horizonte, v. 16, n. 1, p. 165-184, jan./mar. 2011.

PROPER, H. A.; HALPIN, T. A. *Conceptual Schema Optimisation – Database Optimisation before sliding down the Waterfall*. Technical report, Department of Computer Science, University of Queensland, Brisbane, Queensland, Australia, July 1995.

ROWLEY, J. The wisdom hierarchy: representations of the DIKW hierarchy. *Journal of Information Science*, v. 33, n. 2, p. 163-180, 2007.

SEARLE, J. R. Minds, Brains, and Programs. *Behavioral and Brain Sciences*, v. 3, n. 3, p. 417-424, 1980. DOI: 10.1017/S0140525X00005756.

GEITGEY, Adam. *Machine Learning is Fun! – The Book*. 2. ed. [S.l.]: [s.n.], 2019.

JAMES, Gareth; WITTEN, Daniela; HASTIE, Trevor; TIBSHIRANI, Robert. *An Introduction to Statistical Learning: with Applications in R*. 2. ed. New York, NY: Springer, 2021. ISBN 978-1-0716-1417-4; e-book ISBN 978-1-0716-1418-1. DOI: 10.1007/978-1-0716-1418-1.

LI, Zewen; YANG, Wenjie; PENG, Shouheng; LIU, Fan. A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects. *arXiv preprint arXiv:2004.02806*, 2020. Disponível em: <https://arxiv.org/abs/2004.02806>.

SAHARIA, Chitwan; CHAN, William; SAXENA, Saurabh; LIT, Lala; WHANG, Jay; DENTON, Emily; GHASEMIPOUR, Seyed Kamyar Seyed; AYAN, Burcu Karagol; MAHDAVI, S. Sara; GONTIJO-LOPES, Raphael; SALIMANS, Tim; HO, Jonathan; FLEET, David J.; NOROUZI, Mohammad. Photorealistic text-to-image diffusion models with deep language understanding. In: *INTERNATIONAL CONFERENCE ON NEURAL INFORMATION PROCESSING SYSTEMS (NIPS)*, 36., 2022, Red Hook, NY. Proceedings... Red Hook: Curran Associates Inc., 2022. Article 2643, p. 36479–36494.

ENGEL, Jesse; AGRAWAL, Kumar Krishna; CHEN, Shuo; GULRAJANI, Ishaan; DONAHUE, Chris; ROBERTS, Adam. GANSynth: adversarial neural audio synthesis. *ArXiv*, abs/1902.08710, 2019.

HO, Jonathan et al. Imagen video: high definition video generation with diffusion models. *arXiv preprint arXiv:2210.02303*, 2022.

MARANHÃO, Juliano. *Direito de Autor: Desafios Conceituais e a Obra Audiovisual*. *Revista de Direito das Comunicações*, vol. 6/2012, p. 29 – 50.

MAYMON, Giora. Chapter 2 – Some Important Statistical Distributions. In: Maymon, Giora (ed.). *Stochastic Crack Propagation: Essential Practical Aspects*. Amsterdam: Academic Press, 2018. p. 9-18. ISBN 978-0-12-814191-5. DOI: 10.1016/B978-0-12-814191-5.00002-4.

MCAFEE, Andrew. *Generally Faster: The Economic Impact of Generative AI*. Abril 2024. Disponível em: [https://storage.googleapis.com/gweb-uniblog-publish-prod/documents/Generally\\_Faster\\_-\\_The\\_Economic\\_Impact\\_of\\_Generative\\_AI.pdf](https://storage.googleapis.com/gweb-uniblog-publish-prod/documents/Generally_Faster_-_The_Economic_Impact_of_Generative_AI.pdf). Acesso em 27 out. 2025.

MCKINSEY & COMPANY. *The economic potential of generative AI: The next productivity frontier*. Junho 2023. Disponível em: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier?os=vpkn75tqhpmkpsxtq&>. Acesso em 27 out. 2025.

MORAN, Lyle. *54% of in-house legal professionals support generative AI usage*. Thomson Reuters Legal Dive. Junho 2023. Disponível em: <https://www.legaldive.com/news/thomson->

[reuters-generative-ai-legal-use-cases-in-house-legal-corporate-tax/654455/](https://www.reuters.com/legal/generative-ai/legal-use-cases-in-house-legal-corporate-tax/654455/). Acesso em 27 out. 2025.

TYAGI, Kalpana. *Copyright, text & data mining and the innovation dimension of generative AI*. Journal of Intellectual Property Law & Practice, Volume 19, Issue 7, July 2024, Pages 557–570, <https://doi.org/10.1093/jiplp/jpae028>.

UNITED STATES COPYRIGHT OFFICE. *U.S. Copyright Office Fair Use Index*. Agosto de 2025. Disponível em: <https://www.copyright.gov/fair-use/index.html>.

UNITED STATES DISTRICT COURT NORTHERN DISTRICT OF CALIFORNIA. Case 3:24-cv-05417-WHA (Bartz v. Anthropic)

UNITED STATES DISTRICT COURT FOR THE DISTRICT OF DELAWARE. Case No 1:20-cv-613-SB (Thomson Reuters v. Ross Intelligence)

WOLFSON, Stephen. *The Complex World of Style, Copyright, and Generative AI*. Creative Commons. Disponível em: <https://creativecommons.org/2023/03/23/the-complex-world-of-style-copyright-and-generative-ai/>. Acesso em 27 out. 2025.

ZHAO, Jiankun; MENG, Lingwei; DENG, Chengxi; MENG, Helen; WU, Xixin. Defending unauthorized voice cloning with watermark-aware codecs. In: Proceedings of the 26th Annual Conference of the International Speech Communication Association (Interspeech 2025), 17-21 August 2025, Rotterdam, The Netherlands. DOI:10.21437/Interspeech.2025-1993.

XU, L.; XIE, H.; QIN, S-Z. J. Parameter-Efficient Fine-Tuning Methods for Pretrained Language Models: A Critical Review and Assessment. arXiv:2312.12148, 2023.

