



O futuro da pesquisa em inteligência artificial

Anna Helena Reali Costa

Fabio Gagliardi Cozman

resumo

Este artigo discute o futuro da pesquisa em inteligência artificial, argumentando que, em um futuro próximo, técnicas simbólicas continuarão a receber certa atenção, enquanto técnicas de aprendizado de máquina baseadas em dados continuarão a apresentar crescimento explosivo. Modelos obtidos por aprendizado profundo, em especial modelos de linguagem, serão aplicados em variados setores e terão consideráveis melhoras (em eficiência, em interpretação, em desempenho). Ainda no futuro próximo, o debate social sobre inteligência artificial se tornará mais concreto, exigindo esforço de pesquisa sobre legislação, impacto social e mercado de trabalho. Procurando olhar para um futuro mais distante, entendemos que combinações entre métodos simbólicos e baseados em dados receberão maior atenção, enquanto apostas, como computação quântica, poderão levar a novos saltos de qualidade.

Palavras-chave: futuro da IA; aprendizado de máquina; aprendizado profundo; IA e sociedade.

abstract

This paper examines the future of research in artificial intelligence, arguing that, in the near future, symbol-based techniques will continue to receive some attention, while machine learning techniques based on data processing will continue to grow explosively. Models produced through deep learning, in particular language models, will be applied to many sectors and will be significantly enhanced (in efficiency, in interpretability, in performance). Still in the near future, the social debate about artificial intelligence will take a more concrete form, demanding research effort related to regulation, social impact and job markets. Looking into a more distant future, we believe that mixtures between symbol-based and data-based methods will get more attention, while some bets, such as quantum computing, may take us to new performance levels.

Keywords: future of AI; machine learning; deep learning; AI and society.

Vinte anos atrás, prever os rumos da pesquisa em inteligência artificial exigia extrapolar, para um horizonte relativamente curto, um conjunto limitado de subáreas disjuntas. De forma geral, a pesquisa seguiria algumas abordagens consagradas, por exemplo, explorando representações baseadas em lógica, ou aprendizado de máquina inspirado por organismos biológicos. Previsões para um futuro distante eram mais baseadas em ficção científica do que em pesquisa acadêmica. Porém, o horizonte próximo era bem delineado, com algumas linhas de pesquisa oferecendo algum sucesso prático, outras nem tanto.

Por volta de 2007, essa realidade começou a mudar, inicialmente de forma gradual

e recentemente de forma estonteante. Devido ao sucesso prático da tecnologia de IA, cresceu o número de pesquisadores, assim como o volume de investimentos (tanto públicos quanto privados), e sobretudo o número de abordagens e de algoritmos disponíveis na literatura. A previsão do futuro, mesmo o futuro próximo, tornou-se bem mais complexa. Prognósticos sobre IA ficaram mais arriscados, sobretudo em relação a empresas e seus produtos, mas mesmo em relação à pesquisa, que normalmente tem um tempo

ANNA HELENA REALI COSTA é professora titular do Departamento de Engenharia de Computação da Escola Politécnica da USP e coordenadora do Centro de Ciência de Dados (c²d), uma parceria entre o Banco Itaú-Unibanco e a USP.

FABIO G. COZMAN é professor titular do Departamento de Engenharia Mecatrônica da Escola Politécnica da USP e diretor do Center for Artificial Intelligence (C4AI), uma parceria entre a Fapesp, a IBM e a USP.

de maturação mais longo. Neste breve artigo procuramos tatear os caminhos prováveis da IA para o futuro imediato e arriscar algumas poucas previsões para o futuro mais distante. Para tanto, agregamos nossa experiência a uma análise dos principais temas e discussões nas grandes conferências da área, tais como NeurIPS¹, IJCAI² e AAAI³ –, notando que tais conferências hoje atraem tanto atores da academia quanto das grandes empresas de tecnologia. Estas últimas participam não só com interesse em recrutamento, mas também para compartilhar suas mais recentes inovações e para absorver novos resultados da academia.

Certamente há formas diferentes de discutir o futuro da IA, em particular o futuro mais distante. Por exemplo, o recente trabalho de Grace et al. (2024) combina previsões de mais de 2.700 pesquisadores sobre a IA para os próximos 200 anos, focando sobretudo os impactos da tecnologia em setores econômicos e sociais. Aqui fazemos um recorte focado em um prazo menor, com ênfase na pesquisa sobre a tecnologia em si.

ENTENDENDO OS CAMINHOS: *NEATS E SCRUFFIES*

Para entender o futuro, é útil ter um entendimento do passado. A área de IA não é nova, tendo se iniciado por volta de 1950, mas seus objetivos sempre foram

difíceis de especificar. Definir a área é um desafio que recentemente saltou dos livros-textos acadêmicos para o campo da diplomacia internacional. Um órgão que tem se destacado nesse debate é a Unesco, que considera como sistemas de IA aqueles que “têm capacidade de processar dados e informações de uma forma que se assemelha ao comportamento inteligente e, normalmente, inclui aspectos de raciocínio, aprendizagem, percepção, previsão, planejamento ou controle” (Unesco, 2021, artigo 1.2).

Podemos prever, sem chance de errar, um ponto sobre a IA: sua definição continuará a ser discutida infinitamente na academia, na cultura popular, na legislação, na diplomacia internacional. Para nossos propósitos aqui, a lista simples oferecida pela Unesco é bastante adequada. Estamos interessados em sistemas computacionais que representam conhecimento, raciocinam, tomam decisões, aprendem com experiências e interagem com o mundo.

Ao se defrontar com a tarefa de construir um artefato que satisfaça pelo menos a alguns aspectos da definição acima, um pesquisador pode enfatizar objetivos diferentes (Russel; Norvig, 2022). Um possível objetivo é a construção de um artefato que se comporte de maneira “racional” segundo alguma definição formal de racionalidade. Por exemplo, um agente que, ao se defrontar com um problema lógico, saiba operar de forma exata os axiomas da lógica formal. Ou um agente que, ao se defrontar com uma situação de incerteza, saiba calcular probabilidades e comparar o valor esperado de várias alternativas. Um outro possível objetivo é a simulação do comportamento “humano”, ou pelo menos o

1 Ver: <https://nips.cc/>.

2 Ver: <https://www.ijcai.org/>.

3 Ver: <https://aaai.org/conference/aaai/>.

comportamento observado de seres humanos (e, às vezes, de outros animais). Para tanto, podem ser usadas técnicas inspiradas nas propriedades biológicas, ou podem ser combinadas várias técnicas para emular aquilo que é observado: o que importa é imitar, tanto quanto possível, o que vemos no principal sistema que sabemos ser inteligente – o ser humano.

Ocasionalmente pesquisadores cujo foco é a análise formal têm sido chamados de *neats*, enquanto pesquisadores que procuram pragmaticamente combinar variadas inspirações e técnicas têm sido chamados de *scruffies*⁴ (Abelson, 1981). Nas últimas décadas, em alguns momentos a abordagem *neat* predominou, enquanto em outros momentos a abordagem *scruffy* foi vitoriosa.

Por décadas, o principal foco da pesquisa em IA foi a codificação de problemas em linguagens formais, frequentemente baseadas em lógica, e a busca por soluções usando algoritmos gerais (Brachman; Levesque, 2004). Este artigo dará pouco espaço para novos resultados sobre representação formal de conhecimento, raciocínio lógico e racional, tomada de decisões com garantias de desempenho e assim por diante; entretanto, todos esses tópicos fazem hoje parte da agenda de pesquisa de IA e sem dúvida estarão presentes na comunidade de pesquisa em IA durante o futuro próximo. Em particular, os próximos anos certamente trarão melhores algoritmos de geração de planos que nos levem de um estado inicial a um

estado desejado, bem como a melhores linguagens de representação de restrições e de relações entre objetos.

Seja como for, o principal foco da IA passou, nas duas últimas décadas, para a extração de padrões a partir de enormes bases de dados, usando métodos estatísticos ou de inspiração biológica para interpretar observações, medidas, textos e imagens. Embora tanto *neats* quanto *scruffies* estejam bem representados na comunidade de pesquisa, abordagens *scruffy* que se baseiam em dados para reproduzir comportamentos desejados estão obtendo mais vitórias e maior visibilidade.

O APRENDIZADO DE MÁQUINA CONTINUARÁ CRESCENDO FORTEMENTE

A maior disponibilidade de dados na sociedade como um todo e a maior capacidade de distribuí-los e de processá-los computacionalmente levaram, ao longo das duas últimas décadas, a um extraordinário sucesso das técnicas ditas de aprendizado de máquina (Russel; Norvig, 2022).

Sem dúvida, os próximos anos continuarão testemunhando mais aplicações de sucesso de aprendizado de máquina, bem como novas técnicas voltadas a extrair padrões e regras de massas de dados.

Vale a pena diferenciar entre duas classes de métodos de aprendizado de máquina, já que essas duas classes evoluíram e evoluirão de formas diferentes. Para tanto, é preciso compreender alguma terminologia essencial sobre aprendizado de máquina, incluindo termos como aprendizado estatístico, aprendizado profundo, redes neurais e modelos de linguagem.

4 A tradução mais adequada para o português seria, talvez, “empertigados”, para *neats*, e “desgrenhados”, para *scruffies*.

O verbo “aprender” refere-se a um vasto conjunto de fenômenos. É possível aprender lendo uma única vez uma lista de instruções específicas, ou observando um agente resolver um determinado problema algumas vezes, ou coletando uma enorme quantidade de dados, extraindo padrões e reproduzindo esses padrões em ações futuras. Essa última abordagem é a essência do *aprendizado de máquina estatístico* (Hastie; Tibshirani; Friedman, 2009), do qual se extraem padrões de massas de dados tanto por métodos estatísticos consagrados, como regressão ou agrupamento, quanto por métodos relativamente novos, como *boosting* ou SVMs.

Próximo do final da década de 2000 já era claro que as soluções baseadas em aprendizado de máquina alcançavam um desempenho notável, e frequentemente surpreendente, em diversas tarefas anteriormente consideradas desafiadoras. Por exemplo, a tradução automática de textos, por décadas abordada por meio de regras linguísticas, teve enorme evolução quando sua base passou a ser meramente estatística: coletamos sentenças em uma língua e suas traduções em outra língua, e estimamos uma função a partir das sentenças coletadas, que leva as palavras da primeira língua para seus mais prováveis correspondentes na outra língua.

Esse movimento foi muito bem capturado por um artigo influente publicado em 2009, cujo título é muito apropriado: “The unreasonable effectiveness of data”⁵ (Halevy; Norvig; Pereira, 2009). Em suma, os autores desse artigo argumentam que

é viável reproduzir comportamentos aparentemente inteligentes de forma surpreendentemente efetiva simplesmente coletando grandes quantidades de dados e emulando padrões encontrados nesses dados. Note-se que isso foi discutido em 2009! Desde então, a pesquisa em IA baseada em dados só se acelerou e se expandiu.

De certa maneira, o conjunto de técnicas agrupado sob o rótulo de “aprendizado estatístico” compõe um corpo já tradicional de ferramentas que tem encontrado muito sucesso prático. O futuro próximo da IA certamente inclui um continuado e crescente sucesso do aprendizado de máquina estatístico. Pode-se prever que a qualidade de decisões automáticas, por exemplo, em diagnóstico médico, ou em detecção de anomalias em produção industrial, terá significativa melhora nos próximos anos. Haverá um farto espaço para pesquisa em novas técnicas que possam processar mais dados de forma eficiente e sustentável.

Porém, não é no aprendizado de máquina estatístico tradicional que provavelmente serão encontradas novas mudanças de paradigma na pesquisa em IA. Para entender esse ponto, façamos uma retrospectiva sobre o avanço espetacular do *aprendizado profundo*.

Técnicas de aprendizado profundo se popularizaram a partir de 2012, ano em que redes neurais do tipo convolucional venceram uma competição de detecção de objetos em imagens (Krizhevsky; Sutskever; Hinton, 2012). Tais técnicas são baseadas em redes neurais artificiais com muitas camadas, de onde vem o adjetivo “profundo” (o termo original, em inglês, é *deep learning*). Redes neurais artificiais procuram imitar alguns aspectos da orga-

5 Esse título é praticamente impossível de se traduzir com precisão: trata-se de indicar a incrível/surpreendente/espantosa(irracional?) eficácia permitida pelo uso de dados.

nização neuronal observada em organismos biológicos, sendo compostas de camadas que se interconectam, cada camada contendo um conjunto de pequenas unidades de computação simples que são chamadas de neurônios artificiais. A analogia com cérebros biológicos, embora grosseira, é explorada há décadas em IA, em alguns momentos tendo gerado enorme entusiasmo e, em outros, tendo se deparado com ceticismo. Desde o aparecimento do aprendizado profundo, esse ceticismo praticamente desvaneceu e hoje existe farta evidência empírica indicando que redes neurais com muitas camadas têm capacidade de reproduzir padrões que métodos estatísticos mais tradicionais não conseguem aproximar.

De certa forma, a vitória atual do aprendizado profundo é uma extraordinária vitória da filosofia *scruffy*. Redes neurais artificiais são sistemas bastante opacos, cuja análise formal é extremamente complexa; é praticamente impossível explicar de maneira não trivial a razão de uma rede neural artificial produzir uma determinada saída a partir de uma dada entrada.

A efervescência trazida pelo aprendizado profundo, que teve impacto avassalador entre 2012 e 2015, foi seguida de outras pequenas revoluções em rápida sucessão. Em meados da década de 2010 tornou-se comum transformar símbolos em quantidades numéricas, o que melhorou em muito o processamento de linguagem natural⁶. Outra revolução ocorreu em 2017, com o aparecimento de uma organização

particular de neurônios artificiais e camadas que amplia a capacidade de processar dados pelas redes neurais resultantes. Tais redes são os chamados *transformers* (Vaswani et al., 2017) e, em pouco tempo, apareceram os *modelos de linguagem de grande porte* baseados em *transformers*. Um modelo de linguagem recebe símbolos em uma determinada linguagem e produz probabilidades de emissão de outros símbolos (na mesma ou em outra linguagem). Por exemplo, dada a sequência “o aluno entrou na”, um modelo de linguagem produzirá probabilidades para palavras como “sala”, “escola”, “enfermaria” etc. Modelos de linguagem são usados há décadas, porém o primeiro modelo de linguagem de grande porte que fez uso de *transformers* foi o modelo Bert (Devlin et al., 2019), que apresentou um desempenho excepcional em muitas tarefas de processamento de língua natural.

Hoje existem muitos modelos de linguagem de grande porte, indicados pela sigla LLM (do termo em inglês *large language model*). O mais conhecido é provavelmente o modelo GPT, o motor que dá substância ao popular serviço ChatGPT⁷. LLMs são espantosamente proficientes em algumas tarefas, como resumir textos, corrigir a gramática e a sintaxe de um determinado texto, mudar o tom de um texto para públicos específicos e traduzir de um idioma para outro. Ao receber textos adequadamente preparados, certos LLMs conseguem até mesmo emular padrões relativamente sofisticados de

6 Usualmente um mapeamento de símbolos a números é chamado de *embedding* na literatura (Mikolov et al., 2013).

7 Ver: <https://chat.openai.com/>.

raciocínio, embora o princípio básico de seu funcionamento seja bastante simples: em resumo, emitir os símbolos mais prováveis dado o conjunto de símbolos de entrada. Essa variedade de IA opera de forma bem diferente daquilo que imaginamos ser a inteligência humana. Além disso, os LLMs também apresentam problemas. Algumas vezes, LLMs podem incluir dados incorretos declarados como fatos, em um fenômeno chamado de “alucinação” (Maynez et al., 2020). Isso ocorre quando o modelo não possui a compreensão contextual necessária sobre o que foi perguntado. Um LLM não consegue verificar se uma informação está correta ou não e pode mesmo fornecer respostas conflitantes a uma pergunta se esta for feita mais de uma vez. Pesquisadores têm explorado um conjunto de ideias para reduzir o número de alucinações: coleta de mais dados e seleção mais apurada de dados; quantificação da incerteza no processo de geração de palavras; e melhora do processo de estimação de probabilidades de palavras (Kaddour et al., 2020; Ziwei et al., 2023).

O futuro da IA passa por um melhor entendimento dos LLMs e das técnicas a eles associadas. Alguns dos problemas enfrentados por LLMs poderão ser superados com maior coleta de dados. Mas construir um LLM pode ser extremamente custoso tanto em termos computacionais quanto em energia. O significativo gasto de energia para treinamento de LLMs foi inicialmente notado por Strubell, Ganesh e McCallum (2019), um estudo que levou a muitos outros experimentos nos últimos anos. Parece seguro dizer que LLMs têm um custo ambiental, tanto para treinamento quanto para inferência

(Desislavov; Martínez-Plumed; Hernández-Orallo, 2023), mesmo quando examinados a médio prazo (Patterson et al., 2022). Além disso, o custo ambiental em outras dimensões, por exemplo, relativo ao consumo de água, deve ser mais bem equacionado no futuro (IEA, 2024).

Dado o incrível esforço que vemos hoje em torno dos desafios de superar os pontos fracos dos LLMs, podemos esperar para o futuro próximo uma nova geração de modelos, mais eficientes no uso de energia, com maiores garantias de veracidade em suas afirmações e com maior habilidade de se conectarem com programas auxiliares – por exemplo, programas que facilmente resolvam questões matemáticas. Haverá significativo esforço para desenvolver instrumentos teóricos que permitam analisar e compreender modelos baseados em redes neurais profundas e técnicas correlatas, como mecanismos de geração de imagens. Haverá também pesquisas sobre formas de organizar esses modelos de maneira modular, assim combinando competências de modelos diversos. Sobre tudo, veremos intenso esforço de pesquisa centrado em modelos de linguagem que contenham mecanismos de controle e prevenção de alucinações. Isso nos levará aos tão esperados auxiliares artificiais que nos darão apoio e aconselhamento sem absurdas divagações. Muitas empresas têm mostrado que pretendem levar esses avanços para o mercado com celeridade. Com efeito, uma previsão fácil para o futuro próximo é a chegada ao mercado de variados produtos que usam aprendizado de máquina (em particular LLMs). Esse avanço será acompanhado por um melhor entendimento de quais técnicas

são de fato úteis, transcendendo o atual período de agitação intensa.

Seja como for, entre métodos estatísticos e neurais, o futuro próximo da pesquisa em IA será em boa parte voltado a aplicações práticas. Na academia veremos cada vez mais a IA aplicada à compreensão e à antecipação de fenômenos físicos ou sociais, nos mais variados campos do saber. Exemplos simples: a IA pode combinar massas de dados e conhecimento sobre oceanos para prever ressacas e prevenir populações litorâneas; pode ainda combinar milhões de dados do SUS para encontrar a melhor estratégia de vacinação. Veremos também um continuado esforço para tornar menos opacas as decisões tomadas por métodos estatísticos ou neurais. Ou seja, esforços para facilitar a interpretação dessas decisões ou para gerar explicações para essas decisões.

O DEBATE SOBRE IA E SOCIEDADE SE TORNARÁ MAIS CONCRETO

Nos últimos anos a IA tem suscitado diversas preocupações e gerado inúmeros debates. A sociedade tem discutido o impacto da IA no jogo político, bem como examinado a capacidade da IA de produzir falsidades e de infringir a privacidade. Avanços recentes na tecnologia levantam considerações éticas e preocupações sobre o uso responsável da IA e a necessidade de evitar usos maliciosos. Modelos estatísticos e modelos de linguagem podem ser suscetíveis a ataques mal-intencionados que manipulem dados de entrada para alterar decisões. Por serem extremamente complexos, modelos obtidos via aprendi-

zado profundo são difíceis de interpretar, o que dificulta a detecção de vulnerabilidades e de vieses. Experimentos realizados nos últimos anos detectaram decisões de caráter discriminatório em sistemas de IA destinados a aplicações práticas.

Efeitos no mercado de trabalho têm também causado preocupação. O relatório sobre inteligência artificial e empregos, divulgado pela Organização para a Cooperação e Desenvolvimento Econômico, alerta que 27% dos empregos dos seus países membros são de profissões com alto risco de automatização pela IA (OECD, 2023). Recomenda-se ali que nações invistam em capacitação apropriada para trabalhadores.

O futuro próximo verá essas preocupações se transformarem em instrumentos de controle e regulamentação ao redor do mundo. O debate, que até agora se deu mais no campo das ideias, começará a produzir leis e diretrizes concretas. Em toda parte, a legislação será uma ferramenta discutida com maior intensidade, especialmente no que diz respeito à responsabilização por erros realizados pela IA e à quantificação de riscos associados à IA. A considerável discussão em torno do Artificial Intelligence Act aprovado pela União Europeia (2024), que terá efeitos mais adiante, indica o quanto o assunto causa controvérsia.

Entretanto, não é possível prever claramente quais desses esforços terão resultados positivos ou negativos. Não será fácil encontrar o balanço adequado entre incentivo da IA “boa” e controle da IA “ruim”. Os países que adotarem uma abordagem mais cautelosa e ponderada provavelmente alcançarão mais sucesso. Esses movimentos exigirão suporte acadêmico em áreas

como direito, economia e ciências sociais e políticas. Compreender como o ser humano reage a esse novo ambiente ao seu redor, que inclui máquinas com alguma inteligência, e encontrar formas de tornar esse ambiente mais humano e mais justo irão requerer pesquisa séria e continuada.

Uma previsão para o futuro próximo: em breve veremos esforços de empresas, em todos os setores, para avaliar e certificar suas tecnologias de IA como socialmente adequadas. Processos de certificação variados serão tentados: autorregulação, órgãos certificadores, órgãos centrais etc. A regulação via poder público não será o único caminho no futuro.

Outra previsão para o futuro: as nações que optarem por proibir a tecnologia enfrentarão condições econômicas e sociais mais desfavoráveis nos próximos anos do que aquelas que adotarem a IA, mas agirão com cautela.

Como estamos pensando em nações, cumpre aqui anotar um risco trazido pela IA que se superpõe a variados outros fenômenos sociais: haverá um aumento substancial na disparidade entre nações no que diz respeito à IA, e isso não significa apenas um aumento na diferença de infraestrutura computacional relevante ou robôs de serviços ou manufatura, mas sobretudo um aumento em disparidades nas dimensões sociais como educação, trabalho e renda (Alonso et al., 2020). Essas diferenças já se aplicam hoje à pesquisa em IA: enquanto algumas poucas nações estão investindo pesadamente na pesquisa em novas tecnologias, e notadamente em IA, e lideram decisivamente, muitas outras nações permanecem à mercê de avanços produzidos fora de suas fronteiras.

Deve-se notar também uma mudança significativa ocorrida ao longo dos últimos anos e que tem importantes implicações práticas. Até cerca de 2017, a maior parte dos avanços em IA acontecia na academia e em um distribuído conjunto de laboratórios empresariais com longo histórico, por exemplo, em empresas como IBM e Microsoft. Em cerca de cinco anos essa realidade mudou: hoje grande parte da pesquisa de ponta, e em particular uma enorme fração da pesquisa de maior impacto, ocorre em um conjunto de laboratórios de pesquisa de empresas, nas próprias linhas de desenvolvimento de produtos de grandes empresas, e em startups que já nascem com investimentos de porte. Algumas poucas empresas, a grande maioria baseada nos EUA, se projetaram enormemente e hoje lideram essa corrida tecnológica, com OpenAI, Microsoft e Google tendo um notável destaque. Outros países procuram avançar com rapidez, notadamente a China, mas também a Coreia, o Reino Unido e a Alemanha, para citar alguns exemplos. Tudo isso mudou a distribuição da pesquisa em IA em poucos anos, o que dificulta prever o que veremos a seguir.

Encerrando essa seção com uma pequena digressão sobre a posição do Brasil entre as nações, notamos que ele está bem integrado econômica e culturalmente com a comunidade internacional, e a população brasileira é tipicamente receptiva a novas tecnologias; nesse contexto, os avanços que impactam as nações líderes na pesquisa em IA geram impacto semelhante no Brasil. Na academia, o Brasil tem se destacado no cenário mundial de publicações e citações de artigos de IA. Em especial, em 2023, o Brasil figurou em 16º lugar em termos de publicações em IA e em 17º lugar em

termos de citações das publicações em IA, sendo o primeiro da América Latina e Caribe à frente de países como Bélgica, Suécia, Áustria e Dinamarca (OECD.AI, 2024). Têm havido esforços no Brasil para criar vários centros de pesquisa e aumentar o financiamento para inovação acadêmica e empresarial relacionada à IA. O governo brasileiro publicou em 2021 uma estratégia nacional⁸, agora em revisão, para organizar iniciativas relacionadas a apoio, financiamento e regulação. Embora o Brasil já tenha leis em vigor para abordar a proteção de dados e privacidade, há uma considerável discussão em torno de novas legislações destinadas a fortalecer o controle sobre os artefatos de IA. Apesar dos riscos que a IA acarreta, e que devem ser mitigados por uma regulamentação sensata, a tecnologia de IA pode ser um elemento significativo para impulsionar o avanço sustentável do país. Contudo, a questão mais complexa que o país enfrenta atualmente em relação à IA é como lidar com a propagação de desinformação, especialmente para fins políticos. Esse considerável número de desafios exige um investimento sério na pesquisa em IA, bem como incentivos para formação de pessoas com competência nessa área.

OLHANDO ADIANTE

Um LLM capaz de conduzir diálogos interessantes, como o GPT, é gerado a partir de bilhões de documentos coletados na internet. Supondo que colemos

muito mais documentos, mais diálogos, mais argumentos, será que chegaremos a um ponto no qual um modelo de linguagem colossal demonstre real inteligência? A médio prazo, com certeza a pesquisa em IA produzirá modelos de linguagem gigantescos – se eles de fato demonstrarão inteligência em maior profundidade é uma questão em aberto.

Mas o futuro pode nos levar a outro cenário. Pode ser que a capacidade dos LLMs evolua em muitas direções, por exemplo, em menor consumo energético, sem realmente mudar seu patamar de raciocínio. Ou seja, pode ser que as estruturas hoje empregadas não consigam atingir o desempenho que esperamos de uma real inteligência (algo em si bastante difícil de definir...). Como será o futuro da pesquisa nesse caso? Uma possibilidade é que novas estruturas consigam aprender a raciocinar a partir de textos e exemplos. Talvez seja necessário combinar modelos, reproduzindo estruturas sociais, de forma a gerar consensos e resolver problemas. Outra possibilidade é que seja necessário revisitar o uso de conhecimento e raciocínio formal, por exemplo, usando regras lógicas explícitas, para dar o próximo grande salto de desempenho em IA. Vários caminhos se abrem.

Um futuro possível é aquele no qual a IA *neat* reassumirá uma posição de destaque na construção de agentes artificiais inteligentes. Nesse caso, LLMs e seus companheiros passam a prestar serviços de alta qualidade, porém muito especializados. Sistemas práticos poderão usar diferentes técnicas, algumas neurais, outras simbólicas. Outro futuro possível é aquele no qual a conexão entre redes neurais e técnicas simbólicas ocorrerá

8 Ver: https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/arquivos/inteligenciaartificial/ebia-documento_referencia_4-979_2021.pdf.

através de uma simbiose mais complexa, tornando real o que hoje se pesquisa sob o rótulo de “IA neurossimbólica” (D’Avila Garcez; Lamb, 2023).

Seja como for, a IA não é feita só de imitações do comportamento humano; também esperamos frequentemente que aparatos artificiais tenham melhor desempenho que nós próprios em tarefas específicas. Esperamos que um automóvel corra mais do que um ser humano; esperamos que uma inteligência artificial consiga realizar inferências lógicas de forma mais confiável que um ser humano. Nesse sentido, haverá sempre espaço para técnicas simbólicas e abordagens *neats* dentro da IA. O futuro mostrará se esse espaço será pequeno e escondido, ou se ocupará a maior parte do campo.

Uma previsão fácil é que todos esses caminhos serão trilhados por pesquisadores da área de IA. Parece claro que uma união entre as abordagens *neat* e *scruffy* pertencerá ao futuro da IA.

Como a área de IA tem atraído muita atenção e muitos investimentos, mudanças súbitas devem ser esperadas a médio prazo. A previsão é arriscada; infelizmente, ainda não conhecemos nenhuma inteligência que consiga de fato prever o futuro.

Com o tempo, a pesquisa em IA procurará desenvolver auxiliares artificiais ubíquos, capazes de conduzir conversas longas e tomar decisões complexas. Essa IA se tornará mais integrada ao nosso ambiente, mais econômica e sustentável; espera-se que a pesquisa em regulação, ética, trabalho e temas ligados à sociedade também contribua para que cheguemos a uma IA positiva para o ser humano. Ou seja, olhando mais adiante, só se pode especular. Há no momento esperança de que uma união entre computação quântica e IA possa levar a um novo salto de desempenho nesta última, mas a pesquisa é incipiente nessa direção. Outra aposta de longo prazo é o desenvolvimento de conexões diretas entre máquinas e cérebros biológicos, tanto para permitir que pessoas com deficiências específicas possam recuperar certas habilidades, como para expandir a capacidade cognitiva humana. Previsões ainda mais ambiciosas que, por exemplo, se aventuram sobre máquinas superinteligentes que podem causar catástrofes, ou que imaginam artefatos com consciência, estão no momento mais próximas da ficção científica do que da ciência e tecnologia da IA.

REFERÊNCIAS

- ABELSON, R. P. "Constraint, construal, and cognitive science". *Third Annual Conference of the Cognitive Science Society*, 1981, pp. 1-9.
- ALONSO, C. et al. *Will the AI revolution cause a great divergence?* 11/set./2020. Disponível em: <https://www.imf.org/en/Publications/WP/Issues/2020/09/11/Will-the-AI-Revolution-Cause-a-Great-Divergence-49734>.
- BRACHMAN, R. J.; LEVESQUE, H. J. *Knowledge representation and learning*. San Francisco, Morgan Kaufmann, 2004.
- COZMAN, F. G. "No canal da inteligência artificial – Nova temporada de desgrenhados e empertigados". *Estudos Avançados*, v. 35, n. 101, 2021, pp. 7-20.
- D'AVILA GARCEZ, A.; LAMB, L. "Neurosymbolic AI: the 3rd wave". *Artificial Intelligence Review*, 2023.
- DESISLAVOV, R.; MARTÍNEZ-PLUMED, F.; HERNÁNDEZ-ORALLO, J. "Trends in AI inference energy consumption: beyond the performance-vs-parameter laws of deep learning". *Sustainable Computing: Informatics and Systems*, v. 38, 2023. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2210537923000124>.
- DEVLIN, J. et al. "Bert: pre-training of deep bidirectional transformers for language understanding". *Proceedings of NAACL-HLT*, 2019, pp. 4.171-86.
- GRACE, K. et al. "Thousands of AI authors on the future of AI". arXiv 2401.02843v1 (cs.CY), 2024.
- HALEVY, A.; NORVIG, P.; PEREIRA, F. "The unreasonable effectiveness of data". *IEEE Intelligent Systems*, v. 24, n. 2, 2009, pp. 8-12.
- HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. *The elements of statistical learning*. 2ª ed. Nova York, Springer, 2009.
- IEA – International Energy Agency. *Electricity 2024: analysis and forecast to 2026*. Disponível em: iea.org/reports/electricity-2024.
- KADDOUR, J. et al. "Challenges and applications of large language models". arXiv 2307.10169 (cs.CL), 2023.
- KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. "ImageNet classification with deep convolutional neural networks". *Advances in Neural Information Processing Systems*, v. 25, 2012.
- MAYNEZ, J. et al. "On faithfulness and factuality in abstractive summarization". *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020, pp. 1.906-19.
- MIKOLOV, T. et al. "Distributed representations of words and phrases and their compositionality". *Advances in Neural Information Processing Systems*, v. 26, 2013.
- OECD – Organização para a Cooperação e Desenvolvimento Econômico. *OECD employment outlook 2023: artificial intelligence and the labour market*. Paris, OECD Publishing, 2023.
- OECD.AI Policy Observatory. *Visualisations powered by JSI using data from OpenAlex*. Disponível em: www.oecd.ai.
- PATTERSON, D. et al. "The carbon footprint of machine learning training will plateau, then shrink". *Computer*, v. 55, n. 7, jul./2022, pp. 18-28.

- RUSSELL, S. J.; NORVIG, P. *Inteligência artificial – Uma abordagem moderna*. 4ª ed. São Paulo, Ed. GEN LTC, 2022.
- STRUBELL, E.; GANESH, A.; McCALLUM, A. "Energy and policy considerations for deep learning NLP". *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019, pp. 3.645-50.
- UE – União Europeia. *Regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts*. 21/jan./2024. Disponível em: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206>.
- UNESCO – Organização das Nações Unidas para a Educação, a Ciência e a Cultura. *Recomendação sobre a ética da inteligência artificial*. Código do documento: SHS/BIO/PI/2021/1, 2021. Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000381137>.
- VASWANI, A. et al. "Attention is all you need". *Advances in Neural Information Processing Systems*, v. 30, 2017.
- ZIWEI, Ji et al. "Survey of hallucination in natural language generation". *ACM Computing Surveys*, v. 55, n. 12, 2023, pp. 1-38.