

Como pensar a aprendizagem por transferência na era dos modelos fundacionais

A aprendizagem por transferência, ou *Transfer Learning*, consolidou-se como uma das ideias mais poderosas da inteligência artificial moderna. Ela parte de um princípio simples: o conhecimento adquirido em uma tarefa pode ser aproveitado em outra. Se antes essa abordagem representava apenas uma economia de tempo e dados, hoje ela é o pilar central de uma nova era — a dos modelos fundacionais (*foundation models*). Esses modelos são redes neurais de larga escala, treinadas com grandes quantidades e diversidade de dados para aprender representações gerais do mundo, que podem ser adaptadas a diferentes tarefas. Essa mudança, além de técnica, é conceitual: a transferência deixa de ser um ajuste localizado e passa a ser o próprio modo de operação da IA contemporânea.

Os modelos fundacionais, como GPT, BERT, CLIP ou Stable Diffusion, são treinados em larga escala com dados massivos e variados, aprendendo representações gerais do mundo. Em vez de especializarem-se em uma única tarefa, tornam-se modelos-base capazes de sustentar múltiplas aplicações. Assim, tarefas específicas — tradução, classificação, sumarização, análise de sentimento ou geração de imagens — deixam de exigir treinamento do zero. O que ocorre é um processo de adaptação, uma nova camada de aprendizado em cima de um modelo já consolidado.

Nesse cenário, o papel do pesquisador e do desenvolvedor muda profundamente. Não é mais necessário construir uma rede neural do início, ajustar pesos aleatoriamente e esperar convergência após dias de treinamento. Agora, a prática consiste em escolher um modelo pré-treinado, compreender seu comportamento e adaptá-lo ao problema desejado. Essa adaptação pode ocorrer de várias formas: fine-tuning (ajuste completo ou parcial dos pesos), LoRA e Adapters (camadas adicionais de baixa complexidade) ou Prompt Tuning (ajuste de vetores de entrada que guiam o comportamento do modelo sem alterar seus parâmetros internos).

Os grandes repositórios abertos de modelos desempenham papel central nesse novo paradigma. Plataformas como Hugging Face e Jina AI transformaram a distribuição e reutilização de modelos em um ecossistema colaborativo. O Hugging Face, por exemplo, reúne milhares de modelos pré-treinados para visão, linguagem e multimodalidade, permitindo que qualquer pessoa — de estudantes a laboratórios de pesquisa — possa baixar, testar e adaptar arquiteturas de última geração com apenas algumas linhas de código. Esse compartilhamento padronizado de modelos e pesos acelera a inovação científica, pois facilita o acesso a recursos antes restritos a grandes centros de pesquisa.

A Jina AI expande essa lógica ao integrar repositórios de modelos com fluxos de inferência em produção, voltados a aplicações corporativas e sistemas multimodais. Nessa abordagem, modelos especializados — como extratores de embeddings de texto, imagem ou áudio — podem ser facilmente encadeados e expostos como APIs, viabilizando soluções de busca semântica, recuperação aumentada por contexto (RAG) e sistemas de recomendação. Em vez de pensar a transferência como um evento pontual, passa-se a concebê-la como um processo contínuo, onde cada modelo alimenta o próximo com conhecimento acumulado.

Do ponto de vista educacional, compreender essa transição é fundamental. O Transfer Learning deixa de ser uma etapa final de refinamento e passa a ser a estratégia principal de uso de modelos em larga escala. Ele redefine o que significa “treinar um modelo”: não se trata mais de criar do zero, mas de dialogar com um modelo existente, inserindo-lhe novas instruções, domínios ou dados específicos. Essa mentalidade reflete a própria dinâmica do aprendizado humano — ninguém comece sem referências; aprendemos adaptando o que já conhecemos a novos contextos.

Portanto, pensar a aprendizagem por transferência na era dos modelos fundacionais é reconhecer que a fronteira entre “pré-treinar” e “aplicar” se tornou fluida. O conhecimento está encapsulado em modelos compartilhados, acessíveis e continuamente atualizados em repositórios globais. O desafio, hoje, não é mais obter poder computacional suficiente para treinar um modelo, mas entender como reutilizá-lo de forma eficiente, ética e contextualizada. É nesse ponto que a IA contemporânea se encontra: menos centrada em criar estruturas do zero e mais voltada a integrar, adaptar e expandir inteligências já existentes.

Referências

BOMMASANI, Rishi et al. **On the Opportunities and Risks of Foundation Models**. arXiv, 2022. DOI: [10.48550/arXiv.2108.07258](https://doi.org/10.48550/arXiv.2108.07258). Disponível em: [http://arxiv.org/abs/2108.07258](https://arxiv.org/abs/2108.07258).

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep learning**. MIT Press, 2016. Disponível em: <https://www.deeplearningbook.org>

HUGGING FACE TEAM. **Hugging face documentation**. 2025. Disponível em: <https://huggingface.co/docs>.

JINA AI TEAM. **Jina AI documentation**, 2025. Disponível em: <https://docs.jina.ai>.

VASWANI, Ashish; SHAZER, Noam; PARMAR, Niki; USZKOREIT, Jakob; JONES, Llion; GOMEZ, Aidan N.; KAISER, Lukasz; POLOSUKHIN, Illia. **Attention Is All You Need**. arXiv, , 2023. DOI: [10.48550/arXiv.1706.03762](https://doi.org/10.48550/arXiv.1706.03762). Disponível em: [http://arxiv.org/abs/1706.03762](https://arxiv.org/abs/1706.03762).