

Apresentação da obra coletiva

“Proteção de Dados Pessoais: Que futuro Estamos a Construir”

23 de fevereiro de 2026, Auditório António de Almeida Santos, Assembleia da República

Gostaria de começar por agradecer à CNPD o convite para intervir na conferência de 2024 e, agora por esta nova ocasião, também o convite para estar presente na apresentação do livro que reúne as várias intervenções. Recordo-me de que a sessão foi particularmente interessante e continuo a acreditar na importância de promover reflexões conjuntas sobre os impactos da inteligência artificial na sociedade.

Em 2024, na minha intervenção aquando da conferência, falei sobre inteligência artificial generativa e modelos de linguagem de grande escala, centrando-me em três pontos principais: uma explicação, ainda que a um nível relativamente elevado de abstração, de como estes sistemas funcionam; os seus principais potenciais e limitações; e os desafios que colocam em matéria de proteção de dados e, de forma mais ampla, para a sociedade. O meu objetivo principal nessa altura – o qual continua a ser o objetivo principal na maioria das minhas intervenções em eventos de disseminação e reflexão conjunta, semelhantes a este – foi o de clarificar o que esta tecnologia realmente é, antes de abordar as suas aplicações e problemas.

Na altura, sublinhei que estes sistemas não constituem formas verdadeiramente autónomas de inteligência semelhante à dos humanos, mas que por outro lado estes sistemas são também muito diferentes de bases de dados determinísticas ou motores de pesquisa tradicionais, que devolvem respostas previsíveis que foram previamente preparadas por peritos. Na base destes sistemas estão, antes, modelos probabilísticos de grande escala, treinados para prever continuações plausíveis de texto, com base, por um lado, em enormes quantidades de dados recolhidos da Internet e, por outro, em processos de simulação orientados para otimizar aquilo que os seres humanos consideram mais relevante. Estes sistemas vão além de "papagaios" que se limitam a repetir conteúdos na Internet, mas por outro lado a sua fluência também não implica uma compreensão genuína. Este desfazamento entre a "aparência de inteligência" e o mecanismo subjacente é, desde logo, uma fonte central de risco. Salientei igualmente as limitações destes sistemas,

nomeadamente erros factuais apresentados com grande confiança, capacidades de raciocínio limitadas, enviesamentos herdados dos dados de treino (e temos que a Internet reflecte os muitos enviesamentos da sociedade), e a dificuldade – se não mesmo a impossibilidade – de detetar de forma fiável conteúdos gerados por inteligência artificial.

Quase dois anos depois, considero que as observações que fiz continuam válidas, embora o contexto tenha evoluído de forma significativa. Se, na altura, estes sistemas já não eram utilizados apenas por um pequeno grupo de utilizadores pioneiros, isso é ainda mais verdade nos dias de hoje, em que a utilização é verdadeiramente omnipresente. Do ponto de vista técnico, os sistemas atuais são, de facto, mais capazes, embora continuem, por vezes, a falhar de forma espetacular. Tornaram-se também mais multimodais, e estão claramente mais integrados em ferramentas do quotidiano. A minha intervenção anterior incidiu sobretudo sobre os modelos de linguagem em si, mas atualmente estes sistemas são componentes de ecossistemas de software muito mais complexos, que integram múltiplas funcionalidades e capacidades. Processam não apenas texto, mas também imagens e vídeo; estão cada vez mais ligados a sistemas externos, como motores de pesquisa, bases de dados, calendários, e outras ferramentas informáticas; e são utilizados para orquestrar fluxos de trabalho, estando integrados diretamente em contextos profissionais, serviços públicos, e plataformas educativas. Esta integração permitiu mitigar algumas limitações, como o acesso a informação atualizada, mas aumentou também o potencial impacto de eventuais erros, usos indevidos, e fugas de dados.

A sociedade tem-se adaptado à utilização destes sistemas, e as exigências que lhes são colocadas são também cada vez maiores. Para os engenheiros e investigadores que trabalham no desenvolvimento destes sistemas, as preocupações atuais incluem aspetos como o suporte à diversidade de línguas e culturas dos utilizadores, a melhoria da eficiência energética e, por vezes, a garantia de acessibilidade da tecnologia independentemente de constrangimentos comerciais ou geopolíticos – a concentração e a escala dos modelos de inteligência artificial tornaram-se questões relevantes. Tudo isto deve ser assegurado mantendo, simultaneamente, elevados níveis de qualidade nos resultados, o que é extremamente desafiante. Apesar dos enormes impactos que estes sistemas já têm na sociedade, na economia, ou no mercado de trabalho, os mesmos ainda

afirmam coisas absurdas com enorme confiança. Infelizmente, penso que mesmo aquando de uma eventual futura conferência de reflexão, a acontecer daqui a alguns anos, teremos ainda estas limitações nestes sistemas.

Do ponto de vista dos impactos sociais, aquando da minha intervenção anterior, apresentei alguns exemplos focados no domínio da educação. Defendi que a proibição destas ferramentas não era nem realista nem desejável, e que respostas mais eficazes passariam pela promoção da consciência ética, da transparência, e de políticas institucionais claras sobre a utilização de inteligência artificial. Desde 2024, estas tecnologias obviamente avançaram bastante, mas estou moderadamente otimista quanto à forma como a inteligência artificial pode, e está a transformar a educação. As preocupações iniciais estiveram muito centradas na avaliação e na fraude académica. No entanto, observo com agrado que, cada vez mais, colegas meus se preocupam com o impacto nas experiências de aprendizagem, e não apenas nos mecanismos de avaliação. No caso dos estudantes mais jovens, noto também uma tendência para uma utilização mais responsável destas ferramentas. Muitos deles já tiveram contacto com modelos de linguagem ainda no ensino secundário, e parecem ter apreendido, até certo ponto, conhecimentos e experiência sobre o que estas ferramentas podem fazer, e aquilo para que não devem ser utilizadas.

Na minha intervenção anterior, e mais especificamente no que respeita à proteção de dados, manifestei também preocupações relacionadas com os dados pessoais: a recolha em larga escala de dados para treino, a conformidade com os enquadramentos legais existentes em matéria de proteção de dados, e o risco crescente de ataques de "phishing" e engenharia social assistidos por inteligência artificial. Todas estas preocupações se mantêm relevantes. Do ponto de vista técnico, é interessante notar que vários desenvolvimentos recentes conduziram a técnicas mais eficazes para testar se determinados dados foram ou não utilizados no treino de modelos de inteligência artificial. No entanto, fazer com que estes modelos "esqueçam" um determinado pedaço de informação continua a ser problemático, tanto do ponto de vista técnico, como devido à escala global destes sistemas e à opacidade de algumas práticas de treino e implementação. Em termos gerais, existem hoje expectativas mais claras em matéria de transparência, responsabilização, avaliação de riscos e supervisão humana. No entanto,

ao mesmo tempo, os riscos são mais elevados, e fazer o "enforcing" destas expectativas permanece problemático.

Repetindo um pouco o que foi a conclusão da minha intervenção aquando da conferência, acredito que a sensibilização e a educação são absolutamente centrais. Continuo a defender que salvaguardas técnicas e regulação, por si só, não são suficientes. Utilizadores, desenvolvedores, educadores, e também decisores políticos, necessitam de uma compreensão realista do que estes sistemas podem e não podem fazer. O pensamento crítico, a responsabilidade ética, e a literacia digital, são competências essenciais para permitir uma utilização responsável e benéfica destas tecnologias.

Muito obrigado.

Lisboa, 23 de fevereiro de 2026

Professor Associado do Instituto Superior Técnico da Universidade de Lisboa, Professor
Doutor Bruno Martins