

Como todas as tecnologias também a IA provoca **externalidades** que podem ser indesejáveis. Tal como grande parte da indústria gera **poluição** potencialmente mortífera, também a sofisticação e o mau uso das tecnologias digitais, incluindo a IA, pode causar sérios danos à sociedade.

**Mas a Medicina** e cuidados de saúde são campos de aplicação da IA cruciais e dos mais relevantes não só por resultados já alcançados mas, sobretudo, pelo **escrutínio** a que tais sistemas terão de estar sujeitos e à apertada regulamentação já existente ou necessária, o que leva a estritos procedimentos a seguir, limites de ação e obrigatoriedade de uma Ética que tem de abarcar a IA.

Desde sempre advoguei mais ART (como em ARTificial Intelligence) para garantir uma IA para o bem e responsável. **Accountability, Responsibility and Transparency.**

Já acerca de dez anos atrás, vários grupos académicos ou outros investigadores especializados nestas matérias de LNC aplicavam as tecnologias de IA como SVM e depois Deep NN, quer nos USA, quer na Europa. A partir daí, as mega empresas de tecnologia como a Microsoft e a Google, mas não só, anunciaram a sua intenção de investir em IA pois caminhávamos para um mundo de, não só America First, mas de **AI First**.

O rubicão foi ultrapassado quando na Open AI passaram de um Sistema laboratorial como os que estavam em desenvolvimento noutros labs e universidades , o **GPT2** e 3, para um produto

como o CHATGPT que passou de gerador probabilístico de palavras em contexto, que tão úteis seriam para sumarizações ou até tradução automática, para um **oráculo** mais avançado que um pesquisador/browser clássico gerando textos que mais que verosímeis foram tidos por vezes erradamente, como factuais.

Mas independentemente de todo o ruído, uma certeza temos: A do definitivo e importante impacto da IA em geral na chamada **Patologia Computacional**.

Potenciando a bioestatística, a genómica, a proteómica, a bioinformática em geral, a IA tem, para além do desafio de poder integrar dados de várias origens incluindo imagens, sinais e textos de vários tipos, tem de, dizia, resolver e explicitar as questões **éticas** associadas à sua intervenção, tornando a sua prática socialmente aceitável.

Abstenho-me aqui de enumerar algumas das mais importantes e recentes contribuições da IA como a classificação de 71 milhões de mutações genéticas potencialmente causadoras de cancro ou fibrose cística (vejam a AlphaSignal de 25/9) ou o design “atOMICAMENTE preciso” de proteínas eficazes na geração de anticorpos e úteis para vacinas (vejam um dos mais recentes números da Science).

Mas estes desenvolvimentos colocam a hipótese de o processo de tomada da decisão poder vir a não pertencer exclusivamente ao patologista mas a um **painel** incluindo outros especialistas em ciências de dados (especialistas em estatística) e bioinformáticos, o que coloca sérios problemas **éticos**. Além disso, a contínua e massiva transferência de dados sensíveis entre laboratórios e clínicas e grandes bancos de dados coloca o problema da vulnerabilidade e da segurança. A União Europeia produziu em Maio de 2018 o **General Data Protection Regulation** um documento que claramente impõe novas responsabilidades às Organizações que processam dados dos cidadãos da UE para investigação científica.

Se nos centrarmos na **IA Generativa**, claro que com a expansão do uso de LLMs **Grandes Modelos de Linguagem** mal compreendidos podem surgir consequências desastrosas.

Tais modelos tipo Chat GPT como muitos outros, são baseados nos resultados de massivas Redes Neurais Artificiais Profundas, incluindo Transformadores ( que só apareceram em 2017) e um mecanismo de atenção (cada camada da rede concentra-se numa feature, grosso modo), (Attention is all you need” era o título de um paper seminal) treinadas em enormes quantidades de texto disponível em forma digital e que geram novo texto probabilisticamente credível, parecendo respostas humanas corretas. Alguns incluem GANs Generative Adversarial Networks que ajudam a combinar geração com discriminação.

LLMs estão cada vez maiores e mais caros. De 2019 a 2022 cresceram 360 vezes tornando o treino 160 vezes mais caro. LLMs lêem o que levaria 5K anos 24 h/dia o humano a ler.

**Médicos** tem tentado potenciar estas ferramentas para tarefas que envolvem produção de **textos** como apresentações, publicações, documentação, relatórios etc.

Segundo a reputada revista The Lancet no artigo publicado em Junho 2023 Ethics of large language models in medicine and medical research (Hanzhou Li John T Moon Saptarshi Purkayastha Leo Anthony Celi Hari Trivedi Judy W Gichoya) podem-se nomear vários LLMs já para a medicina, como BioGPT (MIT USA), LaMDA (Google, USA), Sparrow (Deepmind AI, UK), Pangu Alpha (Huawei, Shenzhen, China), OPT-IML (Meta, USA), e Megataron Turing MLG (Nvidia, USA).

Por exemplo o LLM da BioGPT foi treinado com os artigos da PubMed e portanto terão uma base científica. Mas não sei se terão uma base ética.

Mas para além de gerar texto novo também podem ajudar a **sumariar** (por exemplo extensos textos e artigos científicos) gerar propostas de projetos ou de candidaturas a bolsas. Podem também gerar **cartas** para os pacientes, traduzindo o jargão médico em linguagem mais acessível, ou ajudar na escrita de artigos gerando outlines iniciais significativos.

Quanto aos problemas **éticos** do uso dos LLMs, eles começam na fase de treino com a potencial ampliação de **tendenciosismo** mais ou menos escondido que pode logo incluir questões de raça, Género, língua, cultura. São facilmente detetadas associações generalistas como as de certas crenças religiosas à violência ou de etnias a violações da lei ou, noutro sentido, de os mais endinheirados serem os mais necessitados de cuidados médicos (precisamente porque já gastaram mais em cuidados médicos).

As implicações éticas do uso da **IA Generativa** na escrita e publicação de textos e na tomada de decisão advém não tanto dos algoritmos desenvolvidos que são matemática e estatisticamente credíveis com taxas de precisão muito elevadas e conhecidas, embora padeçam do efeito da “**black box**”, ou seja opacidade, sendo portanto pouco explicativos

Mas dizia eu que, sobretudo na medicina, as implicações éticas advém muito da falta de **Segurança e Privacidade** dos dados usados, tendo que se Impedir o acesso indevido, e intrusões por Hackers e tornar essencial a necessária confidencialidade desses dados. Impor e seguir protocolos reguladores como o HIPAA (Health Insurance Portability and Accountability Act) dos USA e o General Data Protection Regulation (**GDPR**) produzido na União Europeia em 2018 é crucial. Este GDPR impõe novas responsabilidades nas **Organizações** que processam dados de cidadãos Europeus mesmo para tarefas de investigação científica.

Evitar o **enviesamento** implica constante **curadoria** dos bancos de dados e monitorização dos processos.

Portanto, constante monitorização e auditoria, logo **responsabilização**, deve ser assegurada para obviar uma escondida malevolência social.

A Transparência é mandatária, permitindo que se conheça a garantia da factualidade dos dados e que a informação individual protegida (que pode variar mas, normalmente, inclui raça, género, nome, ou mesmo idade) não está a ser usada.

Demonstrou-se já que não basta **anonimizar**, por exemplo, dados clínicos para garantir a respetiva **privacidade**. Há por vezes conjuntos de outros dados que se **correlacionam**

A **Transparência** dos processos, dando a conhecer inclusivamente a sua falta, e a assunção de **Responsabilidades** pelos erros ou inexatidões nos conteúdos gerados pela IA generativa tem de ser acautelada e melhorada.

**A quem responsabilizar** por eventuais erros do uso dos algoritmos de IA (aqui falando na Medicina)? Quem os usa, ou seja o médico? Quem permite o seu uso, a Organização disponibilizadora dos cuidados? Quem regulamenta, os Ministérios da Saúde, Justiça, parlamentos e governantes? Quem os vende e distribui, Empresas? Quem os desenvolve, Laboratórios de investigação académicos ou não? Quem os testa e certifica? Todos ao seu nível. Muita legislação tem de ser produzida.

Mas a tendência parece imparável: Ainda há pouco vi um Anúncio lançado pela **iLof** empresa do Porto, cujo objetivo é “acelerar a nova era da **medicina personalizada**”

“We're looking for an awesome Senior Research Data Scientist | Machine Learning | Personalized Medicine | Deep Tech

Na **última newsletter da UE AI Act Newsletter #37** de 23/09/23-05/10/23

Os legisladores ainda discutem as questões mais controversas como a vigilância por **reconhecimento facial** biométrico e a questão do **copyright dos materiais usados pelos LLMs** como o ChatGPT. GPT4 treinado em texto e imagens.

Para permitir distinguir textos produzidos por sistemas de IA, a OpenAI lançou a 31 de janeiro deste ano um classificador de textos mas retirou-o em Julho por falta de precisão. Mas há outros produtos como Originality.ai, Writing.com e Copyleaks.

Mesmo a olho se pode detetar a falta de typos, a frequência de certas palavras (em inglês the, it, is) uso constante de frases mais curtas e por vezes alguma repetição de frases, apresentação de factoides com pouca análise.

Também o Joint Research Centre da European Commission escreveu um relatório sobre os **requisitos da cybersegurança**

pois a interação dos LLMs com outros sistemas mesmo informáticos como Sistemas de BD e o próprio Google, aumentará muito o poder e a possibilidade de quebras de segurança.

Realmente acredito que os investigadores de IA poderão avançar cientificamente no sentido certo. Como, por exemplo, incluir em todos os sistemas de IA uma camada de **software lógico**, obrigando a filtrar potenciais decisões ou ações consideradas não éticas. Esta investigação, de uma Ética “**by design**”, é muito atual.

**Finalmente** possivelmente irá acontecer uma combinação inextrincável entre várias ciências e tecnologias para criar sistemas sofisticados, inteligentes e flexíveis, onde **a IA se diluirá**. As chamadas NBIC, combinando Nanotecnologia, Bioengenharia, Informática & Computação (sobretudo IA) e Ciências Cognitivas, contribuirão para tais sistemas Inteligentes no futuro, colocando, inevitavelmente, **questões ainda mais fundamentais** a serem discutidas para o bem da humanidade como a questão dos Cyborgs já em desenvolvimento e teste a nível militar.

**Mensagem final: É crucial manter o human in the loop responsável e responsabilizável!**