



DOS QUATRO TÓPICOS APRESENTADOS, ESCOLHE APENAS UM E ESCREVE UM **ENSAIO FILOSÓFICO** SOBRE O MESMO.

### INDICAÇÕES:

Tens três horas para redigires o teu ensaio, sendo-te concedidos **10 minutos de tolerância no final para gravação do documento** e recolha do mesmo através de **PEN**.

Dos quatro tópicos possíveis, **identifica claramente o tópico escolhido para realizares o teu ensaio**.

**Escreve o teu número secreto** apenas na primeira página do teu ensaio, **no espaço designado**, e **não escrevas** o teu nome ou qualquer tipo de identificação em nenhum lugar, sob pena de desclassificação.

**Vai gravando o teu ensaio**, em intervalos regulares, de 30 em 30 minutos.

**Não alteres a formatação pré-definida** para a realização do ensaio.

Código Secreto
<b>1047</b>

*Tópico escolhido : 4.*

## O abismo no centro da inteligência artificial

### Introdução

Vivemos num mundo dominado por todo o tipo de algoritmos. Estes são uma das principais ferramentas usadas pelos informáticos para o desenvolvimento dos programas altamente complexos que usamos todos os dias. Ao longo da história dos computadores, foram desenvolvidos algoritmos para resolver os problemas de cada época e, juntamente com os próprios computadores, os algoritmos têm aumentado em complexidade e capacidades exponencialmente desde a sua origem, e cada vez mais é encontrada uma solução baseada em algoritmos para problemas que os computadores não eram capazes de resolver previamente.

Desde a chegada do século XXI, vem-se ouvindo cada vez mais sobre o termo “inteligência artificial”. Muitas qualidades são atribuídas ao que chamamos “IA” – qualquer um que preste atenção aos meios de comunicação na década dos 2020 tem ouvido falar de capacidades que seriam quase inacreditáveis há poucos anos atrás, ou de perigos e vantagens destas novas tecnologias – Ora, a IA é um algoritmo. Mais especificamente, um algoritmo extremamente complexo com capacidades extraordinárias, que estende a capacidade dos computadores além do que pensávamos que poderiam fazer há poucos anos atrás. Bem, se há poucos anos tecnologias como o *ChatGPT* e o *Sora* eram inacreditáveis, e hoje fazem parte do nosso dia-a-dia de forma banal, de que poderá ser capaz a IA nas próximas décadas? Será que devemos começar já a pensar nas capacidades que poderemos



ter ao nosso dispor em relativamente pouco tempo, para não cometermos erros graves? Consideremos a seguinte proposta: se um modelo de inteligência artificial pudesse criar uma sociedade sem qualquer tipo de crime e desigualdade, mas ao custo das liberdades individuais, será moralmente correto deixá-lo fazer tal coisa?

Esta pergunta faz-nos pensar, e se nos lembarmos dela quando estamos perante uma situação de injustiça social, fará-nos realmente considerar a opção hipotética. Como desafio ao leitor: imagine um mundo meritocrático, onde não existe o medo de sair à noite ou ir onde quiser. Se quiser ter filhos, imagine estes a viverem em paz, sem stress de serem ultrapassados por pessoas menos competentes apenas devido às suas famílias serem mais ricas. Imagine poder deixar os seus bens num local público sem nenhum medo de ser roubado. Etc..., etc... Se nos forem oferecidas todas estas coisas e mais, “apenas” ao custo das nossas liberdades individuais, como a privacidade ou a liberdade de expressão, devemos aceitar a oferta? Certamente, a filosofia consegue ajudar-nos a tomar esta decisão. Vários pensadores, mesmo que, na grande maioria, não chegaram a viver até à revolução da IA, já ponderaram sobre questões semelhantes ou relevantes relativamente a esta. Mas antes de relacionarmos o pensamento de autores com a inteligência artificial, devemos compreender o que “realmente” se está a passar em relação a esta tecnologia. Serão as suas capacidades tão grandes e o seu potencial tão vasto que podemos realmente questionar-mos com perguntas deste género, ou será que a IA não passará do seu estado atual?

### A inteligência limitada dos algoritmos

A maioria das pessoas que já usou um modelo de linguagem provavelmente já teve o seu momento de surpresa ao observar as suas capacidades pela primeira vez. Realmente, é surpreendente. Afinal, já vemos há décadas os atores de Hollywood a falarem com computadores nos filmes de ficção científica. Ao pensarmos sobre a velocidade do desenvolvimento da tecnologia nos dias de hoje, provavelmente já seríamos capazes de prever as capacidades dos computadores no futuro próximo, mas vê-lo a acontecer diante dos nossos olhos, a responder à pergunta que nós mesmos escolhemos fazer, tem uma qualidade única de surpresa. Quando até os próprios criadores dos algoritmos se surpreendem com a suas capacidades, podemos afirmar com certeza que as capacidades dos modelos de IA recentes são realmente algo extraordinário.



Estabelecendo então o que existe de novo e revolucionário nas novas inteligências artificiais, podemos agora partir para questionar os seus limites. Quão longe poderemos ir? Afinal de contas, os modelos parecem estar a aumentar as suas qualidades mais rápido do que nós aumentamos a capacidade de lhes fornecer poder computacional para continuarem a melhorar, quem diria que não é possível termos o dilema previamente apresentado nas nossas mãos em poucos anos?

Para compreender os limites de um sistema, é necessário entender o seu funcionamento interno. Não sendo demasiadamente técnico, os modelos de linguagem, vídeo, imagem, voz, etc... pertencem a uma classe de algoritmos chamados *transformers*, esta arquitetura permite um aumento substancial das capacidades dos sistemas de aprendizagem tradicionais. O *transformer* mais conhecido de todos é o próprio *GPT*, cuja terceira letra significa precisamente isso. Estes modelos funcionam com base na transformação de informação digital (textos, fotos, sons) em objetos matemáticos, cujas partes individuais se diferenciam umas das outras. Para “ensinar” a si mesmo como escrever, desenhar ou falar, o algoritmo constrói um enorme “dicionário” dos valores matemáticos de cada tipo diferente de elemento que possui, por exemplo, cada palavra tem um conjunto de valores ligados ao seu significado. A partir disto, a IA interioriza os nossos *inputs* e efetua previsões, com base em cálculos estatísticos fundamentados pela informação que possui, para prever qual seria a melhor palavra (ou pixel, ou frequência) para começar o seu *output*, e parte daí em diante, até a matemática lhe dizer que a melhor coisa a fazer é terminar a resposta.

Isto mostra-nos um grande facto, que se irá revelar como um problema: a IA é extremamente dependente nos dados em que foi treinada. Seja um *transformer* altamente avançado ou apenas um software de autocorretor, a IA depende quase sempre dos dados que interiorizou no processo de treino para criar um *output*. Outro facto que podemos encontrar embutido no nosso entendimento da IA é que realmente não está a pensar como um humano, e que não tem nenhum tipo de compreensão interna do que está a fazer, apenas consegue produzir um *output* que corresponde às nossas expectativas. E ainda nos sobra outro problema: a nossa definição do conhecimento assume que existe um *conhecedor*. Até aos dias de hoje, este sujeito conhecedor foi sempre o próprio homem, mas agora que temos acesso a máquinas que aparentam conhecer, abre-se um problema. Como já vimos, a IA não tem essa capacidade, apenas efetua cálculos com valores que não comprehende. Não tem nenhum tipo de memória nem sentidos físicos para conseguir trabalhar com qualquer tipo de informação além dos valores que é criada para processar. Logo, a IA



não conhece nada, e toda a informação que está ao seu dispor baseia-se no que foram escolhidos como dados de treino no momento da sua criação. Resumindo, a IA é altamente dependente nos seus dados de treino e, por consequência nas motivações dos seus criadores quanto à omissão de certos dados. Não é capaz de ter um raciocínio semelhante a um humano, tomando decisões de um modo falível sem qualquer compreensão do que está a fazer, e não possui nenhum conhecimento, não obstante que, aparentemente, saiba tudo sobre o mundo quando lhe fazemos uma pergunta.

### O dilema visto pelos óculos de Foucault

Com uma compreensão plausível dos limites da inteligência artificial, podemos avançar, sem receio, para a fase de análise e resposta ao nosso dilema. Claro que existem vários tipos de respostas, mas irei fundamentar a minha sobre uma combinação de factos científicos com o pensamento de vários autores do século XX. Começando por uma introdução a um autor, Michel Foucault foi um dos pensadores mais influentes do século XX nas áreas da história, sociologia e filosofia. O seu “projeto” baseou-se numa vontade de compreender e, ultimamente, redefinir o nosso entendimento de conceitos fundamentais como o poder e o conhecimento. Na definição tradicional, o poder é algo que é captado e armazenado por um indivíduo ou uma instituição, onde os mais poderosos podem, em geral, controlar os mais fracos por meio do seu poder, de acordo com as suas vontades. Ora, Foucault tem uma visão fundamentalmente diferente do poder, afirma que é uma relação em ação e não um objeto ou ferramenta que permanece nas mãos de uma minoria. Ainda mais fundamentalmente, Foucault diz-nos que o poder tem a capacidade de influenciar o nosso conhecimento, e que o que nós “sabemos” está sujeito a influências exteriores, ou seja, certas maneiras de ver e analisar o mundo não são permanentemente consideradas corretas, mas são um produto das condições sociais do seu tempo.

Voltando à inteligência artificial: se queremos que a IA crie uma sociedade sem crime e desigualdade, vai ter de exercitar algum poder sobre os cidadãos do mundo. Assumindo que o dilema não está a presumir uma tecnologia de controlo de mentes das massas, a IA iria, provavelmente, de tomar o controlo da sociedade por força, mas podemos dar o benefício da dúvida, e dizer que a IA só aplicaria a “justiça” dentro de uma sociedade controlada de indivíduos que escolheram viver nesta. Podemos considerar este problema em forma de camadas, cada uma levantando uma questão filosófica, e sendo a primeira a previamente abordada: “Qual será a extensão



desta sociedade?”, a segunda refere-se ao *como*, “Por que meios irá atingir esta sociedade?”, a terceira refere-se à pergunta de se a IA será realmente capaz de fazer tal coisa “Será que conseguirá fazer isto, mesmo dentro das suas limitações”, e a quarta pergunta “Quem dirá o que é considerada uma sociedade sem desigualdade?”.

Continuando com a segunda questão, Foucault, novamente, tem ideias que nos ajudam a definir os meios de controlo necessários para implementar tal sociedade. O autor define a *disciplina*, que se refere ao uso do poder para controlar aspectos como o espaço, tempo e comportamento dentro das instituições para moldar as ações dos indivíduos de forma sistemática e não forçada. A disciplina estende-se para o *biopoder*, que envolve o uso de métodos ainda mais avançados para obter o controlo biológico sobre uma população. Ao controlar os nascimentos, as mortes, a sexualidade, as doenças, e mais, as instituições são capazes de moldar a sociedade como desejam a um nível ainda mais alto do que com simples táticas disciplinarias. Ora, a nossa IA provavelmente irá usar colecionar dados sobre os sujeitos para criar táticas de disciplina e biopoder especializadas para cada tipo de pessoa, estas táticas poderão ser simultaneamente quase invisíveis à nossa percepção e extremamente eficazes, fazendo-nos agir de modo desejado sem criar conflitos e instabilidade. Relativamente à terceira questão, como já observámos as limitações da IA, sabemos que: (A) depende de dados de treino influenciáveis e (B) não é capaz de conhecer nem raciocinar. Esta pergunta e os problemas que apresenta ligam-se diretamente à quarta e última questão. Se a IA não é capaz de conhecer nem raciocinar por si mesma, é evidente que haverá influência exterior e parcial quanto à definição de igualdade e ao que constitui um crime. Como previamente visto, a IA não decide o significado do que comprehende por si só, sendo dependente da compreensão prévia de conceitos por humanos que explicaram o que cada coisa significa, havendo várias definições em conflito dentro do próprio algoritmo. Concluindo, a IA terá de enfrentar e superar quase todas as suas limitações atuais para conseguir implementar esta sociedade “ideal”, terá também de usar meios possivelmente crueis, mesmo que os cidadãos sejam escolhidos livremente pela sua própria vontade, e, depois disto tudo, terá de escolher entre enfrentar perguntas filosóficas que a humanidade tenta resolver há milhares de anos, ou usar uma definição de um humano com preconceitos e enviesamentos para elaborar um método de conduta de modo a garantir a “justiça”.

**Estará toda a esperança perdida?**



Ainda é possível dar um final benefício da dúvida ao dilema: digamos que temos uma IA capaz de fazer tudo isto por meios com que todos os cidadãos concordaram, e que irá agir com base em definições morais com que todos os cidadãos também concordam. Será assim finalmente justificada?

O pensamento de Foucault é novamente útil aqui. Nas suas obras, analisa uma prisão teórica chamada “o panóptico”, criada por Jeremy Bentham. A prisão baseia-se numa arquitetura em que as celas se situam à volta de uma torre de vigilância, que está desenhada de modo a ser impossível um prisioneiro saber se está atualmente a ser observado ou não. Isto cria uma pressão artificial dentro da mente do prisioneiro, que, estando num estado de superposição entre estar e não estar a ser observado, cede às regras do estabelecimento, e age da forma que lhe foi pedida. Não precisamos de criar uma situação hipotética para comparar uma sociedade a um panóptico, Foucault já o fazia no séc. XX, e hoje podemos fazê-lo, de forma ainda mais aparente, com o regime do partido comunista na China. Neste país, pelo menos nas cidades mais desenvolvidas com os melhores sistemas de vigilância e instituições mais capazes de implementar a disciplina, as pessoas estão sujeitas à vigilância quase constante do governo, havendo câmaras por toda a parte e algoritmos que juntam dados de localização, compras, atividade online, etc... num perfil para cada pessoa. O fim destes perfis é a atribuição de uma pontuação de crédito social a cada indivíduo: quanto mais de acordo o comportamento do indivíduo estiver com o ideal delineado pelo estado, maior será a sua pontuação, e, como consequência, as suas oportunidades dentro do país serão amplificadas, o oposto acontecendo com os que têm pontuações baixas. Se queremos a nossa sociedade “ideal”, a IA terá de implementar uma versão ainda mais extrema deste tipo de sistema, tendo como outras opções apenas um sistema de vigilância que simplesmente prende quem não cumpre o perfil delineado pela IA imediatamente, sem segundas chances, ou algo ainda pior, que mata quem não cumprir as ordens, mantendo os cidadãos sobre um terror absoluto sem escapatória, pior que os piores regimes totalitários.

Certo... mas temos a certeza que a ideia do crédito social não é melhor do que as democracias liberais, que têm estado a priorar em muitas maneiras há décadas? Julgo que a resposta é “Não”. Afinal de contas, se tivermos uma IA tão poderosa, não poderemos simplesmente usá-la como uma ferramenta de ensino e distribuição de informação que torne as escolhas dos eleitores em escolhas mais racionais para o benefício da maioria? Isto teria de ser trabalhado, certamente, mas desenvolver a sociedade em direção a uma democracia mais justa é muito preferível a simplesmente desistir e deixar uma IA implementar uma ditadura sem liberdades



sobre os seus próprios princípios, sobre a qual poderíamos ficar presos e escravizados para sempre como espécie.

## Conclusão

Respondendo ao dilema inicial de uma vez por todas, será que deveríamos deixar uma IA criar uma sociedade sem crime e desigualdade, pelo custo das nossas liberdades individuais? Não! A situação não é apenas impensável atualmente, devido aos limites da inteligência artificial dos dias de hoje e à dependência nas definições subjetivas que iriam, de modo certo, influenciar uma IA supostamente “imparcial”, ao serem definidas previamente por pessoas enviesadas, como também não é preferível ao uso da IA para fins educacionais para as novas gerações, que poderão ser educadas de modo a se sobrepor aos sistemas opressivos das sociedades atuais para implementar novas soluções mais justas. Devemos focar-nos em mudar quem somos, pois, como Foucault nos diz, não somos seres com uma natureza imutável, e podemos mudar para o bem. Por isso, não devemos limitar o nosso potencial com formas de governo opressivas, e devemos manter a nossa liberdade para nos tornarmos os melhores que podemos ser.