

A INTELIGÊNCIA ARTIFICIAL COMO FERRAMENTA PARA ABORDAR A SAÚDE MENTAL

Instituto Superior de Engenharia do Porto
Departamento de Física

Unidade Curricular de Epidemiologia
2023

Prof. Dr. Jorge Mendonça

Prof.^a Dra. Carla Pinto

Prof. Pedro Guimarães

1120198 Alessandro Santos

ÍNDICE

1. INTRODUÇÃO	1
1.1. Contextualização e Motivação	1
1.2. Estrutura do Documento	1
2. SAÚDE MENTAL.....	3
2.1. Determinantes.....	3
2.2. Sintomas	4
2.3. Diagnóstico e Tratamento	5
3. INTELIGÊNCIA ARTIFICIAL	6
3.1. Machine Learning	6
3.1.1. Unsupervised Learning	7
3.1.2. Supervised Learning	7
3.2. Deep Learning.....	8
3.3. Chatbots	9
4. CHATBOTS NA SAÚDE MENTAL.....	10
4.1. Análise Crítica	11
4.1.1. A exclusividade, complexidade e particularidade da saúde mental	11
4.1.2. Questões éticas e legais.....	12
4.1.3. A relação médico-paciente	13
4.1.4. Falta de validação científica.....	13
5. CONCLUSÃO	15
5.1. Apreciação Final	16
REFERÊNCIAS BIBLIOGRÁFICAS	17

1. INTRODUÇÃO

O presente trabalho foi realizado com o propósito de consolidar os conhecimentos adquiridos na unidade curricular de Epidemiologia. Esta disciplina foi lecionada no Mestrado de Engenharia Biomédica (MEBIOM), ministrado pelo Instituto Superior de Engenharia do Porto (ISEP), no Politécnico do Porto (P. PORTO).

O trabalho realizado consistiu no desenvolvimento de uma análise crítica relativa à utilização da inteligência artificial, em particular, *chatbots*, para a triagem, diagnóstico e tratamento de perturbações mentais.

1.1. Contextualização e Motivação

A *World Health Organization* (WHO) define o conceito de saúde como sendo um estado de completo bem-estar físico, mental e social, e não apenas a ausência de uma doença [1].

Das três componentes apresentadas, destaca-se a componente mental que, durante muitos anos, foi completamente desprezada e negligenciada pela sociedade. Tal deve-se à complexidade das doenças mentais, essencialmente pela difícil compreensão, dificuldade do diagnóstico, assim como todo o estigma social em torno destas doenças [2].

Em 2019, estimava-se que os distúrbios mentais afetavam uma em cada oito pessoas [3], sendo que, com a pandemia do COVID-19, o número de pessoas afetadas aumentou entre 24 % a 50 % [4].

Apesar de existirem algumas opções para a prevenção e tratamento, a esmagadora maioria destes casos não é tratada. Algumas das razões passam pelo sentimento de vergonha, desconhecimento ou falta de acesso a cuidados de saúde especializados na vertente mental [3].

Todas estas falhas podem ser, numa primeira fase, colmatadas com a introdução de sistemas inteligentes como, por exemplo, *chatbots*. É possível que estas ferramentas sejam capazes de ajudar na triagem, diagnóstico e tratamento inicial, pelo menos até a intervenção de um profissional de saúde especializado [5].

Desta forma, torna-se relevante e interessante o estudo da utilização da inteligência artificial, como uma ferramenta de primeiro nível para a abordagem de problemas do foro mental.

1.2. Estrutura do Documento

Ao longo deste trabalho é apresentado ao leitor uma análise crítica relativa à utilização dos *chatbots* como uma ferramenta de primeiro nível para a saúde mental.

Numa primeira fase é apresentado o conceito de saúde mental, uma parte integrante da saúde, dando-se particular ênfase aos distúrbios mentais. São apresentados alguns dos determinantes associados, assim como alguns sintomas prevalentes. No fim, é apresentado o processo de diagnóstico e respetivo tratamento.

Em seguida, são apresentados alguns conceitos relacionados com a inteligência artificial importantes para o âmbito deste trabalho. É apresentado o conceito de *machine learning* e os respetivos tipos de algoritmos. Depois, aborda-se o conceito de *deep learning*, em particular, redes neuronais. Por último, são abordados os *chatbots* e o seu funcionamento.

É realizada depois uma ponte entre as duas temáticas apresentadas, a saúde mental e a inteligência artificial. Para tal, é apresentado ao leitor um artigo relativo à utilização de *chatbots* como ferramentas para endereçar a saúde mental. É apresentado um resumo do mesmo, com os múltiplos aspetos positivos e negativos identificados pelos autores. De seguida, é realizada uma revisão e análise crítica do artigo apresentado.

Finalmente, são apresentadas as conclusões deste trabalho. Inicialmente é apresentado um breve resumo de todo o trabalho realizado. E por último, é feita uma apreciação final deste trabalho.

2. SAÚDE MENTAL

A saúde mental é um estado de bem-estar mental que, inclui as componentes emocional, psicológica e social. Trata-se de um direito humano básico e fundamental, inerente a todos os seres humanos, qualquer que seja a sua raça, sexo, etnia, religião ou qualquer outra condição [6].

A saúde mental afeta a forma como as pessoas se sentem, pensam e agem, permitindo que funcionem bem. Assim, uma boa saúde mental permite às pessoas que sejam capazes de lidar com o stress do quotidiano, sendo capazes de realizar corretamente as suas tarefas, tomar decisões, construir relações e contribuir para a sociedade [7] [8].

Existe uma panóplia de distúrbios mentais que podem afetar uma pessoa. Alguns exemplos mais comuns são: depressão; distúrbios alimentares; stress pós-traumático; esquizofrenia; hiperatividade; transtorno obsessivo-compulsivo; transtorno de acumulação; transtorno de personalidade limítrofe (*borderline*); transtorno bipolar; psicoses; entre outros [9] [10].

Como previamente referido, estes distúrbios interferem com o dia-a-dia das pessoas. Existem situações onde essa interferência é leve, não afetando em grande medida uma pessoa, ou seja, não existe nenhuma consequência que influencie o seu estilo de vida. No entanto, o oposto também pode ocorrer. Quando graves, a ausência de diagnóstico e tratamento por parte de um profissional de saúde especializado pode levar a consequências críticas, dando aso a possíveis situações de automutilação ou suicídio [11].

2.1. Determinantes

Quanto se estuda qualquer doença, torna-se importante o estudo dos seus determinantes. Os determinantes de saúde são os fatores que contribuem para o estado de saúde atual de uma pessoa [12].

As origens específicas destes distúrbios ainda não são completamente conhecidas, mas sabe-se de alguns fatores que podem influenciar a saúde mental. Ao longo das nossas vidas, múltiplos determinantes individuais, sociais e estruturais podem surgir e afetar a nossa saúde mental, tanto de uma forma positiva, como negativa [13]. Estes riscos são particularmente críticos quando ocorrem durante períodos mais sensíveis do desenvolvimento humano, mais concretamente, na infância [14].

Fatores psicológicos, biológicos e individuais, como: as competências emocionais; lesões cerebrais; a química cerebral, em particular, a concentração de neurotransmissores; a adoção de um estilo de vida não saudável; eventos ou situações pessoais traumáticas; entre outros; podem tornar as pessoas mais vulneráveis e suscetíveis a problemas de saúde mental [13].

Relativamente aos determinantes sociais e estruturais, a exposição às circunstâncias sociais, económicas, geopolíticas e ambientais, quando desfavoráveis, também aumenta o risco de as pessoas sofrerem perturbações mentais. Exemplos incluem: a pobreza; o desemprego; a exposição à violência; o crime; a guerra; a desigualdade social; a segurança; a exposição a

perigos ambientais; a presença do lixo; as alterações climáticas; a qualidade das escolas, habitações e transportes públicos; o acesso aos serviços de saúde; entre muitos outros [14].

Todos os fatores de risco apresentados podem ser encontrados em diferentes escalas, afetando apenas o próprio indivíduo ou um grupo muito restrito, ou então, afetando comunidades inteiras. Alguns riscos podem ser mitigados pela própria pessoa, enquanto outros necessitam de apoio de outras pessoas.

É importante realçar a complexidade da saúde mental, em particular no que toca à identificação destes determinantes. Isto pelo facto de nem todas as pessoas expostas a estes fatores desenvolverem distúrbios mentais. O oposto também é verdade onde, nem todas as pessoas com perturbações mentais foram expostas a estes fatores. No entanto, é globalmente aceite pela comunidade médica que a interação com estes determinantes afeta de forma negativa a saúde mental das pessoas [13] [14] [15].

2.2. Sintomas

Os problemas de saúde mental são relativamente comuns, surgindo através de, por exemplo, um evento traumático e negativo na vida das pessoas. No entanto, na esmagadora maioria das vezes, estas situações são normalmente temporárias. Quando recorrentes ou contínuas, afetando o quotidiano das pessoas, está-se então perante um possível distúrbio mental, sendo essencial recorrer a um profissional de saúde [11].

Os sintomas originados por problemas mentais podem variar consoante os distúrbios, as circunstâncias e outros fatores. Geralmente afetam o comportamento das pessoas, os seus pensamentos e as suas emoções. Alguns exemplos de sintomas incluem [16] [17]:

- Sentir-se culpado, desconectado, triste ou em baixo;
- Alterações repentinas de humor, apetite ou sono;
- Comportamentos violentos;
- Incapacidade de gerir sentimentos de raiva ou hostilidade;
- Mudanças de comportamento no quotidiano, incluindo na escola ou trabalho;
- Automutilação ou pensamentos suicidas;
- Nervosismo ou ansiedade;
- Apatia, falta de interesse, de energia, cansaço excessivo;
- Dificuldades em relacionar com outras pessoas ou até situações de isolamento;
- Incapacidade de lidar com problemas, desafios ou o stress do dia-a-dia;
- Dificuldade em pensar, concentrar e expressar ideias ou sentimentos;
- Hipersensibilidade, paranoia ou alucinações;
- Entre outros;

Os sintomas apresentados não são fáceis de identificar, em particular, quando ocorrem em alguém desconhecido. Para agravar a situação, na maior parte das vezes não surgem de forma isolada, mas em combinação. Ainda assim, é importante destacar que a presença dos mesmos são um mero indicador, não sendo certo a existência de distúrbios mentais. A confirmação destes problemas deve ser realizada por um profissional de saúde especializado na área da saúde mental.

2.3. Diagnóstico e Tratamento

O diagnóstico de distúrbios mentais é um processo difícil, principalmente pela complexidade das doenças associadas. Não existem testes ou exames concretos que permitam identificar a presença dum distúrbio. Para piorar a situação, a identificação dos sintomas também não é simples. Todos estes fatores contribuem para a dificuldade em obter um diagnóstico correto.

O processo de diagnóstico é normalmente realizado por várias especialidades e compreende vários exames. Inicialmente o paciente pode ser submetido a um conjunto de exames físicos com o propósito de descartar qualquer problema físico que possa explicar os sintomas.

No caso de uma suspeita de um problema de foro mental, um médico especializado realiza um questionário ou entrevista ao paciente, procurando respostas relativamente aos seus pensamentos, estado de espírito, preocupações, comportamentos, histórico familiar, experiências passadas, e qualquer outro tópico que considere relevante para a sua compreensão [18] [19]. Idealmente, estas consultas devem ser realizadas por um médico com especialidade em psiquiatria, uma especialidade que se dedica à prevenção, diagnóstico e tratamento de perturbações mentais, emocionais ou comportamentais [20].

Apesar do estigma social dar muitas vezes a entender que os distúrbios do foro mental são uma escolha ou fraqueza, tal não corresponde à realidade. Não é algo que, se ignorado, desaparece, necessitando do devido tratamento. Existem vários tipos de tratamento, que são avaliados e aplicados caso a caso. O tratamento dos distúrbios mentais depende de vários fatores como o tipo de distúrbio e o grau de gravidade associado [18].

Os principais tipos de tratamento consistem na participação em terapias, tanto individuais, como de grupo. Estes tipos de terapias têm como objetivo explorar os pensamentos, sentimentos e comportamentos do indivíduo, procurando a melhoria do seu bem-estar. Outra abordagem para determinadas situações passa pela criação de um plano de autocuidado, muitas das vezes com seguimento por parte dum técnico. Apoio familiar e de amigos também pode ser uma ferramenta no tratamento. A medicação pode também ser outra abordagem, com particular ênfase nas situações cuja causa é biológica ou quando o paciente se encontra numa fase de recuperação. Casos mais graves podem obrigar à hospitalização da pessoa, onde o paciente pode ser monitorizado e medicado com maior rigor e controlo [21].

De realçar que o tratamento é um processo demorado, e que na maioria das vezes não trata completamente o distúrbio. Mas, a verdade é que pode ajudar, e a falta deste tratamento pode dar origem a situações muito graves, como episódios de violência, automutilação ou suicídio.

3. INTELIGÊNCIA ARTIFICIAL

A Inteligência Artificial (IA), no contexto das Ciências da Computação, é definida como a capacidade de sistemas computacionais serem capazes de simular a inteligência humana. Tal implica que computadores tenham a capacidade de interpretar imagens, tomar decisões, aprender, reconhecer linguagem natural, planejar, identificar padrões, entre inúmeras outras tarefas [22].

O conceito de Inteligência Artificial surgiu em 1950, a partir do trabalho desenvolvido por Alan Turing, que deu origem à publicação do artigo *Computer Machinery and Intelligence* [23]. Ao longo deste artigo, Turing aborda a questão “*can machines think?*” que, em tradução livre, pode ser traduzido para a questão: “será que as máquinas podem pensar?”. Turing considera que os conceitos presentes, o conceito de máquina e pensamento, são difíceis de definir, propondo então um jogo que permite introduzir uma questão similar.

O jogo proposto, inicialmente chamado de Jogo de Imitação e mais tarde preconizado como Teste de Turing, visa colocar um juiz humano a interagir com um sistema e uma pessoa, que se encontram em salas distintas e fechadas. A pessoa deve providenciar respostas verdadeiras e a máquina deve simular respostas humanas, de forma a tentar convencer o juiz de que é, na verdade, um humano. A interação deve ser feita de forma que não haja qualquer pista como, por exemplo, o tom da voz, que permita determinar em que sala está a pessoa e o sistema. O objetivo do juiz é identificar em qual das salas se encontra o humano e o sistema. Para Turing, um sistema que seja capaz de enganar o juiz, levando-o a crer que quando está a comunicar com uma máquina está a interagir com um humano, é um sistema verdadeiramente inteligente [23].

A partir deste trabalho de Turing, a área de IA começou a ser explorada por muitos outros investigadores, dando origem à investigação, ao desenvolvimento e à publicação de inúmeros trabalhos nesta área, introduzindo novos conceitos como *Machine Learning* (ML) e *Deep Learning* (DL).

3.1. Machine Learning

Uma das principais qualidades de um sistema inteligente, é a sua capacidade de aprendizagem. À área responsável pelo estudo e desenvolvimento de métodos que permitam aos sistemas computacionais aprender é chamada de *Machine Learning*. Por outras palavras, ML refere-se à capacidade de sistemas serem capazes de, a partir dos dados disponibilizados, extrair informação e aprender sem a necessidade de um operador humano programar regras de forma explícita [24].

Esta extração de informação é feita através da criação e ajuste automático de modelos matemáticos que, consigam representar da melhor forma possível os dados fornecidos ao sistema [25]. Para tal, é essencial que estes dados sejam tratados de forma a conter única e exclusivamente as informações essenciais, também referidas como características, para a definição de padrões e dos modelos matemáticos [26].

Esta necessidade, aliada ainda ao crescimento nas últimas décadas da digitalização e da internet das coisas (*Internet of Things* - IoT), traduziu-se no aumento exponencial de dados gerados e armazenados, o que, inevitavelmente, possibilitou o crescimento dos sistemas de ML.

Os algoritmos de ML são tradicionalmente divididos em dois grupos [27]: *Unsupervised Learning* e *Supervised Learning*.

3.1.1. Unsupervised Learning

Nos algoritmos baseados em *Unsupervised Learning*, os modelos são criados a partir de dados que não se encontram categorizados. Por outras palavras, ao sistema são disponibilizados dados sem qualquer estrutura ou ordem, cabendo ao próprio a identificação dessa estrutura [27]. Estes algoritmos são particularmente úteis para o reconhecimento de padrões [28].

Normalmente, os algoritmos de *Unsupervised Learning* são utilizados na resolução de problemas de *clustering* ou associação [28].

O *clustering* é um método que consiste em agrupar dados em grupos onde, cada grupo é representado por um conjunto de características semelhantes. O sistema analisa os dados fornecidos, procurando e identificando várias características. Mediante a presença ou ausência de características específicas, o sistema atribui uma categoria a esse conjunto de dados [28]. Alguns exemplos de aplicações de algoritmos de *clustering* na medicina passam pelo estudo da epidemiologia de uma doença [29] ou na divisão de pacientes ou documentos clínicos em grupos [30].

Quanto aos métodos de associação, estes têm como objetivo encontrar relações entre dados. Estas relações são baseadas na sua ocorrência [31]. Um exemplo da utilização destes algoritmos é na prescrição de fármacos onde, quando prescrito um medicamento, podem ser apresentados outros que normalmente são prescritos juntamente com o fármaco inicial [32].

3.1.2. Supervised Learning

Nos algoritmos baseados em *Supervised Learning*, os modelos são criados a partir de dados previamente categorizados, ou seja, os dados fornecidos já se encontram classificados, normalmente por um operador humano. Estes sistemas são capazes de avaliar os dados disponibilizados e comparar o resultado obtido com o resultado esperado. Se necessário, podem depois realizar os ajustes necessários no modelo para que ambos os resultados sejam idênticos. Estes algoritmos são particularmente úteis na resolução de problemas de regressão ou classificação [27].

Os problemas de regressão consistem na previsão de valores contínuos, ou seja, séries temporais [33]. Este tipo de algoritmos visa ajudar a resolução de problemas como, por exemplo, modelos de previsão do comportamento de doenças contagiosas [34].

Relativamente aos problemas de classificação, estes são problemas que requerem a identificação de uma classe a partir dos dados disponibilizados ao modelo. Ou seja, o

resultado é um valor categórico, ou por outras palavras, um valor discreto [27]. Um dos exemplos mais comuns deste tipo de problemas é a classificação de imagens que, como o nome indica, permite que a partir de uma imagem, seja possível categorizar e atribuir-lhe uma determinada classe. Para que esta classificação seja possível, é necessário que o sistema aprenda, geralmente através de casos anteriores, ou seja, imagens previamente classificadas [35].

3.2. Deep Learning

Dentro da área de *Machine Learning*, importa destacar o *Deep Learning*. O DL é um subtópico de ML, responsável pelo estudo de algoritmos de aprendizagem com recurso a Redes Neurais (RN). Por outras palavras, DL pode ser considerada como uma abordagem específica dentro da área de ML que, implica a utilização de redes neurais [36].

As RN são consideradas uma arquitetura que, como o nome indica, pretende simular a estrutura de um cérebro humano, imitando a forma como os neurónios funcionam e comunicam entre si. São compostas por múltiplos nós, chamados de células, divididos em múltiplas camadas. Uma RN possui uma camada de entrada (*input layer*), uma camada de saída (*output layer*) e pelo menos uma camada intermédia (*hidden layers*). Cada uma destas camadas pode ter um número variável de células onde, em cada uma, é realizada uma computação [37]. Na Figura 1 [38] é apresentada a arquitetura de uma RN simples.

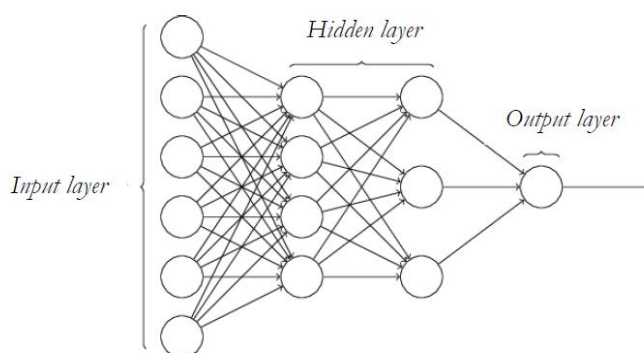


Figura 1 – Arquitetura de uma rede neuronal simples

A computação que ocorre em cada célula consiste no cálculo de um *output* com base nos *inputs*, tendo ainda em consideração os coeficientes, também conhecidos como pesos, que afetam de forma positiva ou negativa o resultado. No final, mediante o resultado obtido, uma função de ativação decide se o processamento deve prosseguir na rede [39]. Este processo encontra-se representado na Figura 2 [39], onde é apresentado um diagrama da computação que ocorre ao nível de uma célula individual. O resultado obtido é também comparado com o resultado esperado, podendo ocorrer o ajuste dos pesos de forma que ambos os resultados sejam iguais [39].

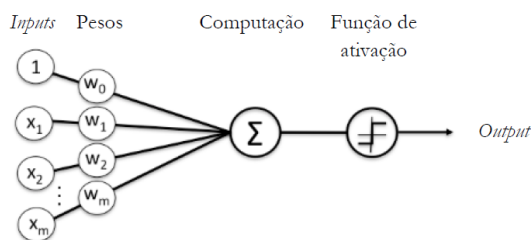


Figura 2 - Computação que ocorre ao nível da célula numa rede neuronal

Assim, uma rede neuronal é, na verdade, uma rede extensa e complexa de computações que ocorrem nas várias células. De forma simples, a rede recebe os dados através da camada de entrada. Posteriormente, os dados são processados nas camadas intermédias, ao nível das células, sendo depois passados para as células seguintes mediante o seu valor. No final, o resultado é apresentado através da camada de saída [40].

Em suma, as redes neuronais destacam-se pela sua capacidade de processar dados de forma paralela, permitindo um processamento eficiente e rápido. São sistemas robustos pela sua capacidade de lidar com dados complexos e incompletos, sendo uma opção viável para problemas relacionados com reconhecimento de padrões e processamento de linguagem natural (PLN).

3.3. Chatbots

Um *chatbot* é um sistema capaz de processar e simular conversas. Estes sistemas podem ser tão simples como um programa rudimentar que responde a questões com uma resposta direta, ou um sistema complexo e sofisticado, como um assistente digital [41].

Estes sistemas têm particular ênfase no tópico de processamento de linguagem natural. O PLN é uma área que se dedica ao estudo e desenvolvimento de algoritmos que tornam as máquinas capazes de interpretar, compreender e manipular a linguagem natural das pessoas. Alguns dos problemas mais comuns que esta área visa resolver são: a tradução de línguas; a extração de informação em texto; o reconhecimento de fala; a identificação de erros gramaticais e de sintaxe; a comunicação entre pessoas com capacidades reduzidas; entre outros [42].

Para o âmbito da presente análise crítica, importa destacar os *chatbots* orientados por dados e preditivos (conversacionais). Estes sistemas são os mais complexos, sofisticados e interativos, sendo normalmente utilizados como assistentes virtuais e digitais [41].

Estes *chatbots* são implementados com recurso a algoritmos de PLN e *Machine Learning*, mais concretamente, *Deep Learning*, ou seja, redes neuronais [43]. Assim, torna-se óbvio que estes sistemas necessitam de ser alimentados com dados para o processo de treino. A qualidade de um *chatbot* está muito dependente da qualidade desses dados, sendo a sua aquisição um desafio, principalmente quando se fala em dados médicos [44].

4. CHATBOTS NA SAÚDE MENTAL

O artigo analisado, de *Boucher et al.* [45], redigido em 2021, faz uma revisão e reflexão da utilização de *chatbots* como uma ferramenta para a intervenção, triagem, diagnóstico e gestão de sintomas e comportamentos associados às doenças mentais.

Numa primeira fase, *Boucher et al.* apresentam o conceito das intervenções digitais em problemas de saúde mental. São apresentados alguns artigos cujos resultados foram positivos, principalmente quando o público alvo eram estudantes do secundário e universitário. Verificou-se que este tipo de ferramentas teve algum sucesso na diminuição parcial de sintomas como a depressão, ansiedade e stress. No que toca a críticas, os autores identificaram que os principais problemas se prendem com a reduzida taxa de utilização e elevada taxa de abandono.

Em seguida, *Boucher et al.* apresenta a utilização da inteligência artificial nas intervenções digitais, fazendo uma breve referência à sua história e evolução. É dada particular relevância aos *chatbots*, apresentando as suas principais funções como o diagnóstico; a disponibilização de conteúdos; e a triagem e gestão de sintomas. São apresentadas algumas evidências relativas à sua utilização como ferramentas para abordar problemas mentais. Estas evidências encontram-se distribuídas em quatro grupos, sendo eles: a aceitabilidade dos *chatbots*; efeitos na adesão dos utilizadores; efeitos na saúde mental dos utilizadores; e, por último, as desvantagens dos *chatbots*.

No que toca à aceitabilidade dos *chatbots*, *Boucher et al.* indicam que o facto de existirem múltiplas soluções e *downloads* de aplicações vocacionadas para a saúde mental, significa que existe um interesse generalizado na área. Os autores referem que no geral, os utilizadores são receptivos a estes sistemas, principalmente pela perceção dada de privacidade; por serem menos críticos e preconceituosos; e pela associação a tecnologia.

Em relação aos efeitos na adesão, *Boucher et al.* voltam a referir que as intervenções digitais apresentam um efeito negativo na adesão dos utilizadores, mas que a utilização de *chatbots* contrariam esta tendência. O motivo apresentado prende-se com o facto de os utilizadores se sentirem obrigados a interagir com o *chatbot*.

Relativamente aos efeitos na saúde mental dos utilizadores, *Boucher et al.* explicam que na data de redacção do artigo, estes sistemas se encontravam a começar a ser explorados, sendo que não existiam resultados concretos sobre os seus efeitos. A maioria dos estudos identificados pelos autores tinham como âmbito a análise da aceitabilidade e usabilidade dos *chatbots* pelos utilizadores.

Por último, no que toca às desvantagens dos *chatbots* identificadas por *Boucher et al.*, estes reconhecem como principal problema a complexidade da linguagem humana. Os autores referem a dificuldade na compreensão de mensagens longas e complexas, assim como de metáforas e hipérboles.

Seguidamente, *Boucher et al.* apresentam um caso de estudo de um *chatbot* comercial chamado Anna, desenvolvido em 2019, pela Happify Health. A sua forma de funcionamento é baseada numa abordagem terapêutica, onde o sistema realiza um conjunto de atividades e

jogos com o utilizador focando-se numa área particular da sua saúde mental. De forma a simular um terapeuta real o melhor possível, os dados utilizados no processo de treino deste *chatbot* são provenientes de profissionais de saúde. Outro ponto também tido em consideração durante o desenvolvimento do sistema, foi a comunicação e relação com o utilizador. O *chatbot* utiliza algumas técnicas que visam criar a noção e ideia de empatia, como a utilização do nome utilizador e perguntas do âmbito pessoal. Os autores realizaram vários inquéritos onde verificam que as pessoas que utilizam o *chatbot* davam respostas mais extensas e completas, assim como mais focadas no assunto em concreto.

Por último, *Boucher et al.* apresentam algumas considerações relevantes e as respetivas conclusões, assim como uma opinião pessoal. Os autores consideram existir muitos preconceitos na utilização da inteligência artificial na saúde mental. Acreditam também ser necessário a realização de mais estudos nesta área, em particular, no que diz respeito aos reais efeitos e resultados na saúde mental. No entanto, os autores consideram que os estudos realizados são positivos e que podem ser percebidos favoravelmente, ajudando a melhorar o envolvimento dos *chatbots* na saúde mental.

4.1. Análise Crítica

No geral, *Boucher et al.* consideram que a utilização de sistemas baseados em inteligência artificial, nomeadamente *chatbots*, são úteis e benéficos na interação com pessoas no âmbito da saúde mental. No entanto, o artigo apresentado deixou de parte alguns pontos fulcrais para esta análise, desde a exclusividade e particularidade das doenças mentais, a questões éticas e legais. Em seguida, são apresentados alguns dos pontos não considerados pelos autores.

4.1.1. A exclusividade, complexidade e particularidade da saúde mental

Inicialmente foi apresentado ao leitor o conceito de saúde mental, abordando-se algumas das doenças associadas, os seus determinantes, assim como o respetivo diagnóstico e tratamento. O principal ponto relativo às doenças mentais que, deve ser retido e considerado para o presente trabalho, é que o diagnóstico deste tipo de perturbações é um processo complexo. Implica normalmente uma ou múltiplas consultas com a pessoa, onde são realizadas várias questões com o propósito de compreender os seus pensamentos, emoções e comportamentos. É importante realçar que não existe nenhum teste ou exame particular para o diagnóstico, sendo este realizado única e exclusivamente com estas sessões [18] [19].

Para complicar ainda mais a situação, é necessário ter em conta que todas as pessoas são diferentes, com as suas próprias características, cultura, ideais e comportamentos, cujas interpretações são subjetivas. Da mesma forma, as causas que originam os distúrbios mentais também são distintas. Tal obriga a que o médico que realize o diagnóstico seja capaz de ajustar as suas questões e atitudes com base no comportamento e respostas da pessoa, que podem ser completamente únicos naquela situação [18].

Tais fatores são importantes e relevantes quando se consideram os *chatbots*, isto porque, como anteriormente indicado, o processo de treino destes sistemas obriga à disponibilização

de dados, ou seja, situações já experienciadas [25]. Assim, estes sistemas podem ser sujeitos a situações únicas, não sendo o sistema capaz de reagir da melhor forma, o que pode ser crítico quando se considera as possíveis consequências dos distúrbios mentais.

4.1.2. Questões éticas e legais

Outra questão pode ser levantada relativamente à origem dos dados utilizados no processo de treino, já que, em teoria, a comunicação entre um paciente e um médico encontra-se sob sigilo médico. Todos os pacientes têm o direito à confidência, isto é, à preservação sigilosa dos factos relacionados com o seu diagnóstico e tratamento. Este é um dos pilares da profissão médica, sendo uma obrigação ética e deontológica do médico [46].

Tratando-se de dados médicos e, eventos e experiências pessoais, para que estes possam ser utilizados para o processo de treino de um sistema, a sua divulgação para estes fins deve ser explicitamente autorizada pelo titular dos dados [47]. Tal, pode implicar a escassez deste tipo de dados, podendo ser um entrave no desenvolvimento destes sistemas.

Deve também ter-se em consideração a nova lei que procura regular a utilização da inteligência artificial na União Europeia. Esta lei foi proposta em 2021 e encontra-se em análise e discussão no parlamento europeu, prevendo-se a sua aprovação e implementação em torno de 2025 [48]. Na forma como está atualmente redigida, a lei da inteligência artificial classifica estes sistemas entre quatro grupos de risco, que implicam a necessidade de cumprir com obrigações e regulamentos distintos. O primeiro grupo corresponde ao do risco inaceitável, cujos sistemas são considerados uma ameaça para as pessoas e são completamente banidos. O segundo grupo é composto pelos sistemas considerados de risco alto. Estes sistemas devem ser avaliados periodicamente durante o seu ciclo de vida. O terceiro grupo inclui todos os sistemas considerados “criativos”, ou seja, que geram dados como texto e imagens. Estes sistemas devem cumprir com regulamentos relacionados com transparência, devendo deixar claro que o conteúdo foi gerado por inteligência artificial. Por último, o quarto grupo é considerado de risco limitado, o menos restrito com apenas alguns requisitos de transparência [49].

Aqui, é necessário ter particular atenção aos sistemas considerados de risco inaceitável, artigo 5 da lei, cujos principais exemplos disponibilizados são [50]:

- *AI systems that deploy harmful manipulative 'subliminal techniques';*
- *AI systems that exploit specific vulnerable groups (physical or mental disability);*
- *AI systems used by public authorities, or on their behalf, for social scoring purposes;*

Se se considerar que os *chatbots* analisados neste trabalho irão interagir com pessoas que sofrem de distúrbios mentais, e que, inclusive, poderão realizar algum aconselhamento, torna-se fácil a inclusão dos *chatbots* neste grupo pelos legisladores. Para suportar esta decisão, é possível considerar casos passados de *chatbots* com comportamentos inapropriados ou ofensivos (derivado dos dados utilizados no treino), como o caso do Tay desenvolvido pela Microsoft e que foi desativado [51] e o GPT-3 que foi corrigido [52].

4.1.3. A relação médico-paciente

Existe ainda outro fator importante que deve ser considerado, a relação médico-paciente. Esta relação é crucial para o sucesso do tratamento durante a prestação de cuidados de saúde, em particular quando se fala da saúde mental [50] [49]. O facto de uma pessoa que sofre de algum distúrbio mental conseguir falar com outro ser humano, abrindo-se e explicando os seus sentimentos, pode muitas das vezes ser uma mais-valia e ajudar no processo de recuperação dessa pessoa [50]. Assim, é fácil de compreender que a interação humana é um fator essencial quando se fala na saúde mental.

No entanto, este fator é desprezado quando se utiliza um *chatbot*, onde o contacto humano não existe. Na atualidade, é relativamente simples e fácil a identificação de um *chatbot*. É necessário que estes evoluam para um nível em que demonstrem um comportamento humano natural, passando despercebidos. Tal pode ser conseguido com, por exemplo, o Teste de Turing, que até hoje nunca foi ultrapassado com sucesso, sendo alguns dos defeitos normalmente apontados o estilo linguístico e os traços emocionais [51]. Aliás, existem todo um conjunto de fatores que destroem esta imersão, tornando evidente que o paciente está a comunicar com uma máquina e não com uma pessoa. Alguns exemplos dos problemas associados à interação com *chatbots* são: a sua insensibilidade emocional; a incapacidade de considerar a especificidade da situação, ou seja, a incapacidade de personalização; o desencorajamento da procura de ajuda especializada; a incapacidade de compreender metáforas; a incapacidade de demonstrar bom senso; entre outros [52].

4.1.4. Falta de validação científica

Outra questão muito importante e, no entanto, quase desprezada pelos autores, é a falta de estudos que apresentem resultados positivos, ou seja, a falta de validação da comunidade científica. Muitos dos estudos existentes não apresentam nenhuma análise ou conclusão que permita inferir sobre os atuais benefícios na saúde mental das pessoas, não apresentando nenhum tipo de comparação com grupos de controlo, ou seja, pessoas que não tenham utilizado sistemas digitais [58] [59] [60].

Outro ponto identificado importante de referir, é a possibilidade de muitos dos estudos analisados não serem completamente isentos, tendo sido realizados por pessoas com algum tipo de relação com o sistema avaliado [60] [61]. Aliás, este facto é comprovável também no artigo analisado onde, os autores do mesmo trabalham na empresa que desenvolveu o *chatbot* apresentado no caso prático.

Os pontos apresentados são relevantes na discussão do papel dos *chatbots* na interação com pessoas com possíveis distúrbios mentais, no entanto, os mesmos foram deixados de parte na análise realizada por *Boucher et al.*. Existem sérias limitações no desenvolvimento e utilização destes sistemas, a começar pela disponibilidade dos dados, importantíssima quando se fala em sistemas baseados em inteligência artificial. É necessário ainda considerar a possível limitação legal imposta pela lei da inteligência artificial, pelo menos se aprovada em parlamento, que poderá limitar o desenvolvimento destes sistemas. Os riscos associados

também são muitos graves, podendo afetar de forma séria o bem estar físico e mental de um paciente. Tal é particularmente crítico quando se consideram sistemas com capacidade aprendizagem, que podem assimilar conhecimentos ou experiências que não sejam as mais corretas e, conseqüentemente, aconselhar ou dizer algo de errado aos paciente.

Assim, considerando todos os pontos identificados, torna-se perceptível que os supostos benefícios possam não compensar os riscos, pelo menos numa situação de comunicação bidirecional.

5. CONCLUSÃO

Ao longo deste trabalho é apresentado ao leitor um resumo crítico sobre a utilização da inteligência artificial, com recurso a *chatbots*, para a intervenção em situações de possíveis distúrbios mentais.

Numa primeira fase, é realizado uma introdução à saúde mental, abordando-se algumas das doenças associadas, os seus determinantes e o respetivo diagnóstico e tratamento. De seguida, é apresentado o conceito de inteligência artificial, em particular as redes neuronais e os *chatbots*. A compreensão do seu funcionamento é importante para o desenvolvimento de uma opinião formada relativa à utilização da inteligência artificial como ferramenta para intervenção em situações de perturbações mentais. Depois, foi apresentado um resumo do trabalho desenvolvido por *Boucher et al.*, apresentando-se os principais pontos do artigo e terminando-se com uma análise crítica do mesmo.

Verificou-se que existem algumas limitações técnicas que podem limitar, senão mesmo inviabilizar o desenvolvimento deste tipo de soluções. A possível escassez de dados, com a agravante do aparecimento da lei da inteligência artificial, podem ser fatores determinantes no futuro destes sistemas. Para piorar a situação, é necessário ainda considerar todas as diferentes situações que podem afetar a saúde mental, sendo impossível o desenvolvimento de um sistema capaz de lidar com todas as situações. A remoção da possível relação paciente-médico, fator determinante para o sucesso não só do diagnóstico, mas também do tratamento, também deve ser algo a considerar. Todos estes pontos são fulcrais para um sistema desta natureza, sendo que as conclusões extraídas desta análise tendem a descartar estes sistemas como uma solução viável.

No entanto, poderia ser interessante uma maior aposta em sistemas que interagissem de forma unidirecional, ou seja, sem existir qualquer comunicação do sistema para o paciente, o que permitiria mitigar os possíveis riscos clínicos associados.

Por exemplo, poderia ser uma mais valia um sistema que, com base no comportamento digital de um utilizador como, o seu histórico de pesquisas ou as suas conversas, fosse capaz de alertar os serviços sociais ou de saúde em caso de uma suspeita de perturbação mental grave. Outro exemplo que poderia ser interessante, seria um sistema capaz de analisar respostas de um paciente a um questionário e, com base no seu conteúdo, realizar uma classificação para servir para a priorização de casos.

Ambos os exemplos, apesar de levantarem questões no que toca à privacidade de dados, não implicariam qualquer tipo de risco clínico para o paciente. Esta interação seria completamente transparente para o mesmo, servindo única e exclusivamente, para alertar e priorizar casos urgentes. Não haveria a hipótese de qualquer aconselhamento ou comunicação indevida e errónea ao paciente. E todos os pacientes seriam atendidos pelo profissional de saúde, independentemente do resultado obtido, não se procurando a sua substituição.

Assim, concluímos que sistemas baseados em inteligência artificial, como os *chatbots*, não aparentam ser bons candidatos como ferramentas para interagirem com pessoas com

possíveis perturbações mentais. Considerando-se uma melhor alternativa a utilização de soluções do género como ferramentas de apoio, sem qualquer tipo de comunicação com o paciente.

5.1. Apreciação Final

O trabalho realizado ao longo deste relatório foi relevante para o desenvolvimento pessoal e profissional do autor. Permitiu a análise de duas áreas interessantes mas pouco exploradas no Mestrado de Engenharia Biomédica, a saúde mental e a inteligência artificial, permitindo alargar o conhecimento do autor, abrindo novas possibilidades. No geral, todo o trabalho desenvolvido foi positivo.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] WHO, "Constitution of the World Health Organization," 22 Junho 1946. [Online].
- [2] Jonas Hill Hospital & Clinic, "If Health is Wealth, Why Do So Many Ignore Mental Health," 8 Dezembro 2022. [Online].
- [3] WHO, "Mental disorders," 8 Junho 2022. [Online].
- [4] S. Nochaiwong, C. Ruengorn, K. Thavorn, B. Hutton, R. Awiphan, C. Phosuya, Y. Ruanta, N. Wongpakaran e T. Wongpakaran, "Global prevalence of mental health issues among the general population during the coronavirus disease-2019 pandemic: a systematic review and meta-analysis," *Scientific Reports*, 2021.
- [5] E. Boucher, N. Harake, H. Ward, S. Stoeckl, J. Vargas, J. Minkel, A. Parks e R. Zilca, "Artificially intelligent chatbots in digital mental health interventions: a review," *Expert Review of Medical Devices*, 2021.
- [6] Nações Unidas, "Declaração Universal dos Direitos Humanos," 1948. [Online].
- [7] WHO, "Mental health," 17 Junho 2022. [Online].
- [8] CDC, "About Mental Health," Centers for Disease Control and Prevention, Abril 2023. [Online].
- [9] S. Dattani, "How are mental illnesses defined?," Our World in Data, 26 Maio 2023. [Online].
- [10] A. Felman, R. Tee-Melegrito e M. White, "Types of mental health disorders," Medical News Today, 23 Dezembro 2022. [Online].
- [11] SAMHSA, "Mental Health Conditions," Substance Abuse and Mental Health Services Administration, 24 Abril 2023. [Online].
- [12] DGS, "Determinantes de Saúde," Plano Nacional de Saúde, 2021. [Online].
- [13] WHO, "Determinants of mental health," 17 Junho 2022. [Online].
- [14] M. Alegria, A. NeMoyer, I. Falgas, Y. Wang e K. Alvarez, "Social Determinants of Mental Health: Where We Are and Where We Need to Go," *Current Psychiatry Reports*, 2018.
- [15] M. Compton e R. Shim, "The Social Determinants of Mental Health," *Focus*, 2015.
- [16] Mayo Clinic, "Mental illness - Symptoms," 13 Dezembro 2022. [Online].
- [17] APA, "Warning Signs of Mental Illness," American Psychiatric Association, 2022. [Online].
- [18] A. Morin e S. Grans, "Getting a Mental Health Diagnosis," Very Well Mind, 2023.
- [19] HealthDirect, "How is mental illness diagnosed?," HealthDirect, 2023.
- [20] CUF, "Psiquiatria," CUF, 2023.
- [21] MHA, "Mental Health Treatments," Mental Health America, 2023.
- [22] J. McCarthy, "What is Artificial Intelligence?," Stanford University, Stanford, 2007.
- [23] A. Turing, "Computing Machinery and Intelligence," *Mind*, pp. 433-460, 1950.
- [24] University of California, "What Is Machine Learning (ML)?," University of California, Berkeley, 2020.
- [25] IBM, "What is machine learning?," IBM, New York, 2023.
- [26] L. Cruz, "Diferença entre Dados e Informação," ExpertDigital, 2022.
- [27] I. Salian, "SuperVize Me: What's the Difference Between Supervised, Unsupervised, Semi-Supervised and Reinforcement Learning?," NVIDIA, 2018.
- [28] javaTpoint, "Unsupervised Machine Learning,," javaTpoint, 2023. [Online].

- [29] C. Nicholson, L. Beattie, M. Beattie, T. Razzaghi e S. Chen, "A machine learning and clustering-based approach for county-level COVID-19 analysis," *PLOS One*, 27 Abril 2022.
- [30] T. Ruzgas e M. Lukauskas, "Data clustering and its applications in medicine," *New Trends in Mathematical Sciences*, pp. 67-70, 2022.
- [31] U. Kant, "Comparing Association Rule Mining with other similar methods," Medium, 2022.
- [32] A. Yoosofan, F. Ghajar, S. Syat, S. Hamidi e F. Mahini, "Identifying Association Rules among Drugs in Prescription of a Single Drugstore Using Apriori Method," *Intelligent Information Management*, pp. 253-259, 2015.
- [33] javaTpoint, "Supervised Machine Learning," javaTpoint, 2023. [Online].
- [34] T. Zhou, "A Linear Regression Prediction Model of Infectious Disease Spread Based on Baidu Migration and Effective Distance," *Comput Math Methods Med*, 2022.
- [35] G. Boesch, "A Complete Guide to Image Classification in 2023," Visio.ai, 2023.
- [36] L. Hardesty, "Explained: Neural networks," MIT News, 2017.
- [37] IBM, "What are neural networks?," IBM, New York, 2023.
- [38] P. Priya, "Classification of Neural Network," Educba, Mumbai, 2022.
- [39] C. V. Nicholson, "A Beginner's Guide to Neural Networks and Deep Learning," A.I. Wiki, 2023.
- [40] T. Shin, "A Beginner-Friendly Explanation of How Neural Networks Work," Towards Data Science, 2020.
- [41] Oracle, "What is a chatbot?," OCI, 2023.
- [42] P. Nadkarni, L. Machado e W. Chapman, "Natural language processing: an introduction," *Journal of the American Medical Informatics Association*, pp. 544-551, 2011.
- [43] IBM, "How chatbots work," IBM, 2023.
- [44] S. Javaid, "7 Chatbot Training Data Preparation Best Practices in 2023," AI Multiple, 2023.
- [45] E. M. Boucher, N. R. Harake, H. E. Ward, S. E. Stoeckl, J. Vargas, J. Minkel, A. C. Parks e R. Zilca, "Artificially intelligent chatbots in digital mental health interventions: a review," *Expert Review of Medical Devices*, pp. 37-49, 2021.
- [46] Diário da República, "Regulamento n.º 228/2019, de 15 de março," Imprensa Nacional-Casa da Moeda, Lisboa, 2019.
- [47] Comissão Europeia, "O que são dados pessoais?," Comissão Europeia, 2018.
- [48] BBC, "AI: EU agrees landmark deal on regulation of artificial intelligence," BBC, London, 2023.
- [49] European Parliament, "EU AI Act: first regulation on artificial intelligence," European Parliament, 14 Julho 2023. [Online].
- [50] European Parliament, "Artificial intelligence act," Junho 2023. [Online].
- [51] P. Lee, "Microsoft: Learning from Tay's introduction," 25 Março 2016. [Online].
- [52] E. Strickland, "OpenAI's GPT-3 Speaks! (Kindly Disregard Toxic Language) Ready or not, powerful text-generation AI will soon be assuming roles of customer service reps and video game characters," 1 Fevereiro 2021. [Online].
- [53] V. Harbishettar, K. R. Krishna, P. Srinivasa e M. Gowda, "The enigma of doctor-patient relationship," *Indian J Psychiatry*, p. S776-S781, 2019.
- [54] B. K. Plesnicar, "Relationship between a Doctor and a Patient with Mental Disorder," *Medicine Law and Society*, pp. 11-19, 2016.
- [55] D. Umberson e J. K. Montez, "Social Relationships and Health: A Flashpoint for Health Policy," *J Health Soc Behav*, pp. S54-S66, 2011.

- [56] C. Jones e B. Bergen, "Does GPT-4 Pass the Turing Test?," *Artificial Intelligence*, 2023.
- [57] S. Coghlan, K. Leins, S. Sheldrick, M. Cheong, P. Gooding e S. D'Alfonso, "To chat or bot to chat: Ethical issues with using chatbots in mental health," *Sage Journal*, 2023.
- [58] B. N. Rudd e R. S. Beidas, "Digital Mental Health: The Answer to the Global Mental Health Crisis?," *JMIR Ment Health*, 2020.
- [59] A. A. Abd-Alrazaq, A. Rababeh, M. Alajlani, B. M. Bewick e M. Househ, "Effectiveness and Safety of Using Chatbots to Improve Mental Health: Systematic Review and Meta-Analysis," *J Med Internet Res*, 2020.
- [60] K. Weisel, L. Fuhrmann, M. Berking, H. Baumeister, P. Cuijpers e D. Ebert, "Standalone smartphone apps for mental health-a systematic review and meta-analysis.," *NPJ Digital Medicine*, 2019.
- [61] L. Donkin, H. Christensen, S. Naismith, B. Neal, I. Hickie e N. Glozier, "A Systematic Review of the Impact of Adherence on the Effectiveness of e-Therapies," *Journal of Medical Internet Research home page*, 2011.