

PROCESSAMENTO DE LINGUAGEM NATURAL UTILIZANDO REDES NEURAIS PARA IDENTIFICAR INTERLOCUTORES EM ARQUIVOS DE ÁUDIO

GUIDO, Jeferson Eduardo.¹
GIORDANI, Fernando.²

RESUMO

O presente artigo relata como foi o desenvolvimento de uma Rede Neural capaz de identificar interlocutores em arquivos de áudio. O Objetivo principal foi identificar esses interlocutores com precisão em arquivos de áudio, bem como apresentar a precisão de um algoritmo tão complexo quanto uma inteligência artificial. As redes neurais, por sua vez, se tratam de um modelo matemático capaz de simular os neurônios de seres inteligentes e tomar decisões por si só. A construção desta rede neural feedforward se deu pela utilização de um modelo sequencial (Sequential Model) e entropia cruzada categórica (Categorical crossentropy). Com isso em mente, é possível fazer paralelos com a fonética forense o que aumentaria muito a produtividade dessa área de atuação bem como auxiliaria em momentos em que seja necessário realizar a detecção de múltiplos usuários ao mesmo tempo. Dadas as particularidades desse algoritmo junto a fonética forense, o ramo jurídico também pode usufruir muito dessa tecnologia por ser uma ferramenta que identifica interlocutores, por exemplo o ramo jurídico conseguirá de forma mais facilitada identificar suspeitos de dado arquivo de áudio.

PALAVRAS-CHAVE: Redes Neurais, Fonética Forense, Inteligência artificial, IA no judiciário.

ABSTRACT

This paper reports on the development of a Neural Network capable of identifying callers in audio files. The main objective was to accurately identify these interlocutors in audio files, as well as to present the accuracy of an algorithm as complex as an artificial intelligence. Neural networks, in turn, are a mathematical model capable of simulating the neurons of intelligent beings and making decisions on their own. This feedforward neural network was built using a Sequential Model and Categorical crossentropy. With this in mind, it is possible to draw parallels with forensic phonetics, which would greatly increase the productivity of this area of expertise as well as help in times when it is necessary to perform the detection of multiple users at the same time. Given the particularities of this algorithm along with forensic phonetics, the legal branch can also benefit a lot from this technology as it is a tool that identifies interlocutors, for example the legal branch will be able to more easily identify suspects from a given audio file.

KEYWORDS: Neural Networks, Forensic Phonetics, Artificial Intelligence, IA in the judiciary.

1 INTRODUÇÃO

Visivelmente, a tecnologia faz-se cada dia mais presente no cotidiano e em atividades que são executadas pelos seres humanos, desde situações corriqueiras, a saber, entretenimento ou atividades profissionais, até algumas mais específicas, a exemplo de atividades científicas ou educacionais (LAURINDO; SOUZA, 2017).

O progresso recente na compreensão da teoria da Inteligência Artificial fez com que ela caminhasse lado a lado com os avanços das tecnologias, tornando-a mais integrada e capaz de se relacionar com outras áreas do conhecimento (RUSSELL e NORVIG, 2013).

¹Acadêmico do curso de Engenharia de Software do Centro Universitário FAG.
E-mail: JefersonEduardoGuido@gmail.com.

²Mestre em Engenharia Agrícola e Cientista da Computação, docente do Centro Universitário FAG.
E-mail: fernando.giordani10@gmail.com.

Utilizam-se redes neurais artificiais, que são técnicas computacionais de modelos matemáticos inspirados em uma estrutura neural de organismos inteligentes e que adquirem conhecimento por meio da experiência. Uma rede neural pode ter inúmeras camadas de unidades de processamento; já o cérebro humano, pode ter bilhões de neurônios.

A propriedade mais importante das redes neurais é a habilidade de aprender de seu ambiente e, com isso, melhorar seu desempenho. Isso é feito por meio de um processo iterativo de ajustes aplicados a seus pesos: o treinamento. O aprendizado ocorre quando a rede neural atinge uma solução generalizada para uma classe de problemas. Com isso em mente, podemos utilizar redes neurais para analisar áudios com diversos interlocutores, em que um algoritmo pode determinar várias diferenças entre os interlocutores, bem como o espectrograma da melodia, da tonalidade, do contraste e também os coeficientes cepstrais da frequência da melodia (MFCCs) de uma série temporal e uma transformação de Fourier de curto prazo (STFT) para utilizar no espectrograma.

Cerca de metade dos tribunais brasileiros possuem projetos de inteligência artificial operantes ou em desenvolvimento – em sua maioria, a partir do trabalho feito por equipes próprias. É o que aponta o relatório da pesquisa Tecnologia Aplicada à Gestão dos Conflitos no Âmbito do Poder Judiciário Brasileiro, produzido pelo Centro de Inovação, Administração e Pesquisa do Judiciário da Fundação Getúlio Vargas (CIAPJ/FGV). (JUDICIÁRIO, 2021).

Nesse contexto, o presente trabalho pretende analisar e descrever a aplicação de redes neurais artificiais na automatização e reconhecimento de vozes de diferentes interlocutores.

Este projeto está estruturado da seguinte maneira: na primeira parte, apresenta-se a introdução, o assunto e o tema do projeto, o problema de pesquisa, os objetivos e a justificativa; na segunda parte, está exposto o referencial teórico, no qual se contextualiza a inteligência artificial, as redes neurais e suas importâncias para o âmbito jurídico e forense; a terceira parte discorre sobre os métodos utilizados nesta pesquisa; são descritos os procedimentos metodológicos da pesquisa na quarta parte e, por fim, apresentados o cronograma e as referências bibliográficas utilizadas.

2 FUNDAMENTAÇÃO TEÓRICA

O presente capítulo tem o objetivo de ilustrar os conceitos sobre a Inteligência Artificial e Redes Neurais acerca do processamento de linguagem natural. Aborda-se e evidencia-se a qualidade de um produto ao aplicar uma Rede Neural na identificação de

pessoas em conversas de áudio e como o uso dessa tecnologia será implementado com os tipos de Redes Neurais existentes atualmente.

2.1 INTELIGÊNCIA ARTIFICIAL: CONCEITOS, HISTÓRIA E EVOLUÇÃO

A concepção de Inteligência Artificial vem sendo abordada e trabalhada há décadas por pesquisadores que buscavam, por meio de lógica matemática e linguagem, mostrar como uma máquina poderia aprender e simular o que foi aprendido (RUSSELL e NORVIG, 2013).

Segundo Barreto (2001), com o surgimento dos primeiros computadores, originou-se uma admiração muito grande pelas inúmeras possibilidades que as máquinas tinham, fazendo com que elas ficassem conhecidas como cérebros eletrônicos, que tinham o objetivo e a capacidade de resolver problemas que até o momento eram reservados aos homens. No entanto, a ideia de que as máquinas poderiam resolver problemas que eram identificados por homens trouxe desconfiança. A solução encontrada pelos fabricantes foi o investimento em propaganda, salientando que eles serviam para auxiliar várias tarefas correntes, como contabilidade, controle de estoque, folha de pagamento e administração de empresas (BARRETO, 2001).

Durante a década de 50 até a década de 70, a Inteligência Artificial passou por um momento de construção, ou seja, de conceituar e descobrir o que poderia ser feito em diferentes teorias e abordagens. Nesse período, ela foi muito usada para operações aritméticas, a fim de demonstrar que as máquinas podiam realizar funções pequenas e exemplificar do que elas eram capazes (RUSSELL e NORVIG, 2013).

Ainda para Russell e Norvig (2013), o aperfeiçoamento e o entendimento das técnicas e da linguagem de IA foram sendo abordados por meio de experimentos e trabalhos dos pesquisadores; com o passar do tempo, ganharam mais prestígio e espaço no mundo dos computadores. Porém, sua evolução foi interrompida, pois, naquela época, pela falta de recursos físicos das máquinas, era impossível continuar e obter sucesso. A partir desse momento, essa tecnologia parte para a resolução de problemas que haviam surgido durante a primeira década, o que ficou conhecido como a abordagem de métodos fracos, que buscava reunir passos elementares de raciocínio para encontrar soluções completas, mas que não permitiam aumento de escala para instâncias de problemas grandes ou difíceis.

Com o tempo, o número de computadores vendidos foi crescendo, com a ideia de que essas eram máquinas que auxiliavam os homens. Na mesma época, um grupo de

pesquisadores continuou a pesquisar a inteligência dos computadores e, pouco a pouco, começou a obter sucesso em suas pesquisas em Inteligência Artificial, trazendo resultados significativos. Todavia, foi com o anúncio de um projeto japonês, que pretendia construir uma nova geração de computadores inteligentes, que o cidadão comum começou a ouvir falar sobre Inteligência Artificial. Com o projeto, vários governos interessados no assunto começaram a financiar essas ideias, o que impulsionou a área de IA (BARRETO, 2001).

Apesar de a Inteligência Artificial ter crescido nesse período, logo após o ano de 1988, a IA sofreu uma fase chamada de “inverno da IA”, momento em que muitas empresas que buscavam as soluções de problemas extravagantes deixaram de investir, o que ocasionou o esquecimento (RUSSELL e NORVIG, 2013).

Posteriormente a esse momento, houve uma revolução no trabalho dessa tecnologia, tanto no conteúdo quanto na metodologia. Diante disso, começaram a ser usadas teorias existentes como base, em vez de propor teorias inteiramente novas. A partir desse momento, a Inteligência Artificial começou a fundamentar as afirmações em teoremas rigorosos ou na evidência experimental rígida (RUSSELL e NORVIG, 2013).

O progresso recente na compreensão da teoria da Inteligência Artificial fez com que ela caminhasse lado a lado com os avanços das tecnologias, o que a tornou mais integrada e capaz de se relacionar com outras áreas do conhecimento (RUSSELL e NORVIG, 2013).

A Inteligência Artificial tem várias definições, de maneira que não se pode definir diretamente “O que é IA?”. Em geral, o melhor conceito dado à Inteligência Artificial é a racionalidade, ou seja, um sistema é racional quando ele faz a coisa certa, segundo o que foi ensinado (RUSSELL e NORVIG, 2013).

Segundo Poole e Mackworth (2010), a Inteligência Artificial é o campo que estuda a síntese e análise dos sistemas de agentes computacionais que agem de maneira inteligente; isso é entendido como algo que atua em um ambiente e que realiza alguma coisa. Um agente pode agir de maneira inteligente quando ele é apropriado para suas circunstâncias e objetivos, é flexível para mudar ambientes e objetivos, aprende com suas experiências que lhe foram ensinadas e, a partir dessas experiências, toma decisões.

De acordo com Fernandes (2003), a palavra inteligência vem do latim *inter* (entre) e *legere* (escolher), ou seja, significa que é aquilo que permite que o ser humano escolha entre uma coisa e outra. A inteligência seria a habilidade de realizar, de forma eficiente, uma determinada tarefa. Já a palavra artificial, vem do latim *artificiale*, que significa algo não natural. Dessa forma, a Inteligência Artificial é um tipo de inteligência produzida pelo homem para dotar as máquinas de algum tipo de habilidade que simula a inteligência do homem.

As definições de Inteligência Artificial variam de acordo com duas características: o pensamento, o raciocínio e o comportamento. Na Tabela 1, são apresentados os principais conceitos da Inteligência Artificial:

Tabela 1 - Definições de IA

Pensando como um humano	Pensando racionalmente
“O novo e interessante esforço para fazer os computadores pensarem (...) <i>máquinas com mentes</i>, no sentido total e literal.” (HAUGELAND, 1985 <i>apud</i> RUSSELL e NORVIG, 2013, p. 4). “[Automatização de] atividades que associamos ao pensamento humano, atividades como a tomada de decisões, a resolução de problemas, o aprendizado...” (BELLMAN, 1978 <i>apud</i> RUSSELL e NORVIG, 2013, p. 4).	“O estudo das faculdades mentais pelo uso de modelos computacionais.” (CHARNIAK e MCDERMOTT, 1985 <i>apud</i> RUSSELL e NORVIG, 2013, p. 4). “O estudo das computações que tornam possível perceber, raciocinar e agir.” (WINSTON, 1992 <i>apud</i> RUSSELL e NORVIG, 2013, p. 4).
Agindo como seres humanos	Agindo racionalmente
“A arte de criar máquinas que executam funções que exigem inteligência quando executados por pessoas.” (KURZWEIL, 1978 <i>apud</i> RUSSELL e NORVIG, 2013, p. 4). “O estudo de como os computadores podem fazer tarefas que hoje são melhores desempenhadas pelas pessoas.” (RICH and KNIGHT, 1991 <i>apud</i> RUSSELL; NORVIG, 2013, p. 4).	“Inteligência Computacional é o estudo do projeto de agentes inteligentes.” (POOLE <i>et al.</i>, 1998 <i>apud</i> RUSSELL e NORVIG, 2013, p. 4). “AI... está relacionada a um desempenho inteligente de artefatos.” (NILSSON, 1998 <i>apud</i> RUSSELL e NORVIG, 2013, p. 4).

Fonte: Russell e Norvig (2013).

2.2 REDES NEURAIS ARTIFICIAIS: CONCEITOS, HISTÓRIA E EVOLUÇÃO

Antes de explicar o que é uma rede neural artificial, precisamos entender melhor o que originou o termo redes neurais artificiais; para isso, precisamos entender algo sobre nós mesmos, que torna cada ser humano consciente e capaz de aprender: o neurônio.

Segundo Ana Maria, no final do século XIX e início do século XX, o italiano Camillo Golgi e o espanhol Santiago Ramón y Cajal (Prêmio Nobel 1906 da Medicina) constataram “estarem os neurônios compondo uma rede de fibras interconectadas”.

Cajal estudou o tecido em embriões de galinha e concluiu que essas células são distintas, têm redes simples que se tornam mais complexas com o tempo (as características que permeiam um neurônio são: plasticidade, crescimento e desenvolvimento). Ana Maria

aponta que essas redes seriam as unidades funcionais dos neurônios. O neurônio possui três partes básicas:

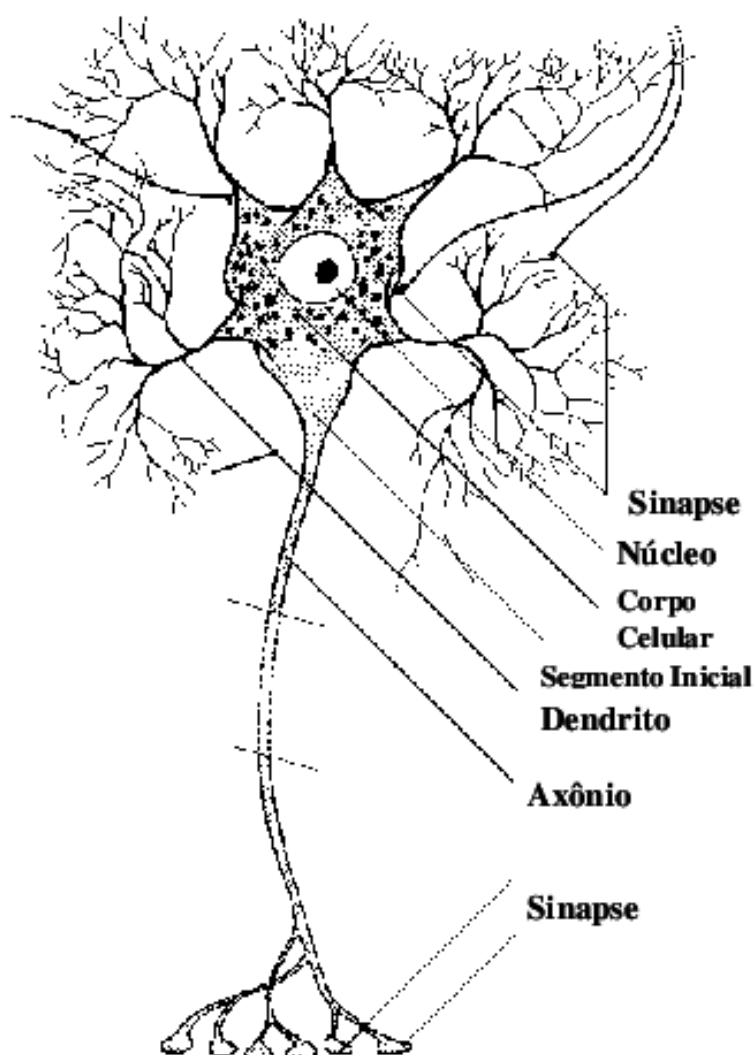
- Um corpo celular de onde fica localizado seu núcleo, grande parte das organelas celulares e de onde partem os prolongamentos dessa célula;
- Prolongamentos ramificados (dendritos), que partem do corpo celular e que garantem a recepção dos estímulos pelo corpo celular.
- E uma raiz principal (o axônio), que é o prolongamento que garante a condução do impulso nervoso. Cada neurônio possui apenas um axônio, o que é, geralmente, mais longo que os dendritos. O axônio possui um isolamento elétrico chamado bainha de mielina.

Os dendritos são múltiplos e o axônio único.

Moreira (2017) evidencia que os “neurônios” são unidades morfofuncionais do sistema nervoso, que recebem informações (sinais elétricos) de outros neurônios e de neuroreceptores especializados, integrando tais informações em suas “áreas operacionais” e encaminhando-as, ao final do processo, na forma de “uma mensagem”.

Redes Neurais Artificiais são técnicas computacionais representadas por um modelo matemático inspirado na estrutura neural de organismos inteligentes e que possuem a capacidade de adquirir conhecimento por meio da experiência. Porém, ainda existe uma grande diferença quando se trata de poder de processamento, por exemplo, uma grande rede neural artificial pode ter centenas ou milhares de unidades de processamento; já o cérebro de um mamífero, pode ter muitos bilhões de neurônios, aponta Edberto Ferneda.

Figura 1: Exemplo de neurônio.



Fonte: https://www.researchgate.net/figure/Figura-32-Exemplo-de-Neuronio-Natural_fig2_228588719.

O sistema nervoso é formado por um conjunto extremamente complexo de células, os neurônios. Eles têm um papel essencial na determinação do funcionamento, comportamento do corpo humano e do raciocínio. Os neurônios são formados pelos dendritos, que são um conjunto de terminais de entrada, pelo corpo central e pelos axônios, que são longos terminais de saída.

Os neurônios se comunicam por meio de sinapses, que são a região onde dois neurônios entram em contato e impulsos nervosos são transmitidos entre eles. Esses impulsos elétricos são recebidos por um neurônio A, são processados e atingem um dado limiar de ação; o neurônio A dispara um novo impulso, produzindo uma substância neurotransmissora que flui do corpo celular para o axônio, que pode estar conectado a um dendrito de um outro neurônio B. O neurotransmissor pode diminuir ou aumentar a polaridade da membrana pós-sináptica, inibindo ou excitando a geração dos pulsos no neurônio B. Esse processo depende de vários fatores, como a geometria da sinapse e o tipo de neurotransmissor. (ANDRÉ CARLOS, 2021)

Segundo André Carlos, em média, cada neurônio forma entre mil e dez mil sinapses. O cérebro humano possui cerca de 10 E11 neurônios e o número de sinapses é de mais de 10 E14, possibilitando a formação de redes muito complexas.

Um histórico resumido sobre Redes Neurais Artificiais deve começar por três das mais importantes publicações iniciais, desenvolvidas por: McCulloch e Pitts (1943), Hebb (1949) e Rosenblatt (1958). Essas publicações introduziram o primeiro modelo de redes neurais simulando “máquinas”, o modelo básico de rede de auto-organização e o modelo Perceptron de aprendizado supervisionado, respectivamente (ANDRÉ CARLOS, 2021).

Alguns históricos sobre a área geralmente deixam de fora os anos 60 e 70 e simplesmente apontam um novo começo da área com a publicação dos trabalhos de Hopfield (1982), relatando a utilização de redes simétricas para otimização e de Rumelhart, Hinton e Williams que introduziram o poderoso método Backpropagation.

Entretanto, para se ter um histórico completo, é preciso citar alguns pesquisadores que realizaram, nos anos 60 e 70, trabalhos importantes sobre modelos de redes neurais em visão, memória, controle e auto-organização, como: Amari, Anderson, Cooper, Cowan, Fukushima, Grossberg, Kohonen, von der Malsburg, Werbos e Widrow.

Em 1959, Bernard Widrow e Marcian Hoff desenvolveram alguns modelos denominados “ADALINE” e “MADALINE”. Uma rede neural utilizando o modelo ADALINE foi desenvolvida para reconhecer os padrões binários de uma transmissão de linha telefônica, de modo que pudesse prever o próximo bit. A rede neural criada com o modelo MADALINE foi a primeira aplicada a um problema do mundo real, usando um filtro adaptativo que elimina ecos nas linhas telefônicas, aponta Widrow.

Uma das propriedades mais importantes de uma rede neural artificial é a capacidade de aprender por intermédio de exemplos e fazer inferências sobre o que aprendeu, melhorando gradativamente o seu desempenho. As redes neurais utilizam um algoritmo de aprendizagem, cuja tarefa é ajustar os pesos de suas conexões (BRAGA; CARVALHO; LUDEMIR, 2000).

2.3 A INTELIGÊNCIA ARTIFICIAL NA ADVOCACIA

Em 2018, uma startup denominada LawGreex desafiou 20 advogados para uma “batalha” entre homem e máquina, que consistia em examinar riscos contidos em cinco contratos de confidencialidade (NDAs – non-disclosure agreements), como aponta João Ozorio.

O resultado já era esperado. A IA obteve precisão de 94%, enquanto a média dos 20 advogados, conquistou 85%. A inteligência artificial executou a tarefa em inacreditáveis 26 segundos, sendo que o advogado mais rápido do grupo levou 51 minutos (EDUARDA, 2019).

Segundo Maria Eduarda, a entrada da inteligência artificial no ambiente jurídico proporciona redução dos trabalhos mecânicos realizados por humanos, ou seja, os processos manuais repetitivos que são executados hoje, no futuro, podem ser eliminados pela automatização.

A firma de consultoria McKinsey estima que 23% do trabalho jurídico pode ser automatizado. Porém, outros aspectos importantes do cotidiano dos advogados não sofrerão ataques das máquinas (OZORIO, 2018).

O interessante é que o robô não existe para tirar o desempenho humano, mas sim para automatização dos serviços com processos repetitivos; o ser humano está aí para o que ele faz de melhor, que é “pensar”, afirma Sheila Aparecida.

Com a evolução da inteligência artificial, desde 2001 até os dias atuais, a empresa IBM usa uma plataforma chamada Watson de serviços cognitivos da IBM para negócios. A cognição consiste no processo que a mente humana utiliza para adquirir conhecimento a partir de informações recebidas (EDUARDA, 2019).

A inteligência artificial já está transformando a rotina de escritórios de advocacia e departamentos jurídicos no País; seu papel já vai muito além da automação (BERETTA, 2019). Segundo Beretta (2019), com o aprendizado da máquina, processamento de linguagem natural e outras soluções cada vez mais avançadas, hoje, a tecnologia tem sido uma importante aliada dos advogados nos processos de análises de contratos, gerenciamento de conteúdo e demais tarefas que exigem trabalho manual.

Nils J. Nilsson (1998, p.1) explica que a Inteligência Artificial, entre vários objetivos, busca o desenvolvimento de máquinas com comportamento inteligente, ou seja, que possam perceber, raciocinar, aprender, comunicar e agir em ambientes complexos tão bem como humanos podem fazer, ou possivelmente melhor.

Bragança (2019, p. 68) afirma que, por meio do mecanismo *machine learning*, “a máquina aprende com as informações colocadas por humanos e a partir daí desenvolve sua própria capacidade cognitiva e decisória”. Esclarece, ademais, que a máquina pode assimilar de forma autônoma e racional algumas situações.

Cerca de metade dos tribunais brasileiros possuem projetos de inteligência artificial operantes ou em desenvolvimento – em sua maioria, a partir do trabalho feito por equipes próprias (STJ, 2021).

Segundo o judiciário do STJ, foi feito um levantamento, no qual atualmente existem 64 projetos de inteligência artificial em 47 tribunais, além da plataforma operada pelo Conselho Nacional de Justiça (CNJ).

2.4 FONÉTICA FORENSE

À fonética, tem-se atribuído o papel de estudar os sons da linguagem humana do ponto de vista material ou físico, descrevendo detalhadamente como eles são produzidos e quais são os seus efeitos acústicos (PEREIRA *et al.*, 2015).

A fonética, como disciplina que estuda o som linguístico em seu aspecto material, pode ser subdividida em três áreas bem definidas: a fonética articulatória, a fonética acústica e a fonética auditiva. A primeira descreve e classifica os sons da fala de acordo com sua articulação no aparelho fonador. À segunda cabe o estudo das propriedades físicas dos sons linguísticos e do percurso que as ondas sonoras trilharam para chegar aos ouvidos do ouvinte. Já a terceira ocupa-se da maneira como os sons da fala são captados pelo aparelho auditivo e interpretados pelo cérebro humano. (PEREIRA *et al.*, 2015, p. 11 e 23.).

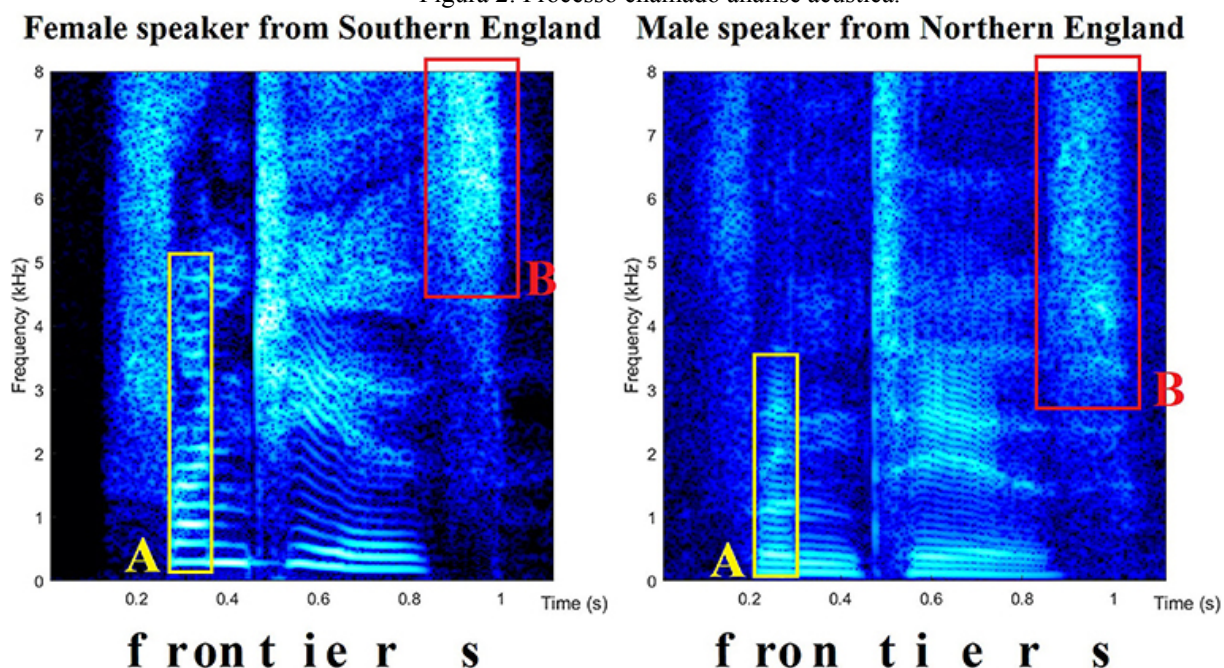
Dessa forma, podemos afirmar que a fonética forense é uma ciência forense e também um procedimento de perícia, que atua na transcrição de áudios e conversas telefônicas para a identificação de traços característicos da fala de determinado indivíduo. Com o procedimento, podem ser verificadas a região e a condição social do falante, seu estado emocional no momento da fala e outras informações diversas. É um tipo de perícia que exige bastante tempo de trabalho, pois nem sempre o material de trabalho, ou seja, o áudio, está em boas condições.

Gully (2022) aponta que as vozes são analisadas por experts treinados em linguística, em várias línguas. Linguistas examinam vários aspectos de uma voz em uma conversação para que, então, eles possam comparar as similaridades juntos da polícia.

Segundo Perito de áudio (2021), a perícia de áudio pode ser realizada em diversos idiomas, tais como transcrever o áudio original em inglês, passando-o para um papel ou equivalente. Alguns deles são altamente sensíveis e exigem alto nível de sigilo, tais como de gravações ou transcrições para registros de reuniões de temas patentários, entre outros.

Os analistas de voz quebram as falas em vários pedaços para escutar com cuidado cada som específico na conversação. Eles também podem examinar as vozes por meio de um software de computador que “observa” as imagens do som das vozes, por meio de um gráfico denominado espectrograma (GULLY, 2022).

Figura 2: Processo chamado análise acústica.



Fonte: Francesco Aletta, 2021.

A imagem apresenta o espectrograma de duas pessoas falando a palavra “fronteiras”. Na imagem, é mostrado o som como energia, em que podemos ter a frequência e o tempo da fala. Existem várias diferenças entre os dois locutores, por exemplo: a caixa A apresenta o som das letras “ro” e “s”, mas são diferentes para cada locutor. Computadores podem ser utilizados para comparar vozes automaticamente, considerando um processo chamado reconhecimento de voz automática (GULLY, 2022).

Para se utilizar do reconhecimento de voz automática, é preciso fazer a análise por computadores e ter uma grande base de dados para determinar a similaridade e o quão distintas duas vozes podem ser.

3 METODOLOGIA

Tomando como ponto de partida os objetivos do estudo, adotou-se uma pesquisa de natureza aplicada, tendo em vista que se concentra em torno dos problemas que permeiam as atividades de organizações, instituições, grupos ou atores sociais. Dessa forma, possui objetivos como: elaborar diagnósticos, identificação de problemas e busca de soluções,

respondendo a uma demanda elaborada por clientes, atores sociais ou instituições (THIOLLENT, 2018).

Devido à peculiaridade e falta de acervo sobre os assuntos tratados, optou-se por uma pesquisa exploratória, pois, como defende Casarin e Casarin (2012, p. 40), trata-se de “uma pesquisa acerca de um tema pouco explorado ou, então, sobre um assunto já conhecido, visto sob nova perspectiva, e que servirá como base para pesquisas posteriores”. Colaborando com esse viés, Gil (2017) afirma que pesquisas que assumem papel exploratório acabam por envolver levantamento bibliográfico e, geralmente, estudos de caso.

As buscas foram feitas por meio do Google Acadêmico, bases de dados, como da Scielo e outras universidades, além de sites de tecnologia, inovação e livros sobre o assunto. Foram utilizadas palavras-chaves sobre o tema para encontrar os conteúdos e agregá-los ao trabalho, as quais são: rastreabilidade, sistema, informação; também se buscou selecionar materiais recentes para aprofundar o estudo.

A pesquisa contou com a etapa de levantamento bibliográfico, que é quando se faz uso de livros, artigos e *sites* com relação ao tema, para discorrer sobre o estudo, a fim de cumprir os objetivos propostos. Na sequência, foram executadas as etapas de coleta de requisitos de sistema, prototipagem das funções e desenvolvimento e testes do sistema. No levantamento de requisitos, foram coletadas e analisadas as funcionalidades necessárias para atender à proposta e à adequação da aplicação.

Por seguinte, na etapa de desenvolvimento tem como premissa implementar as funcionalidades mapeadas nos processos anteriores em uma aplicação de terminal, tanto no que tange o *back-end*, em que se localiza a maioria das regras de negócio, quanto em relação à parte *front-end*, que é responsável pela visualização final de dados por parte do usuário.

O sistema irá possuir o algoritmo de uma rede neural que seja capaz de analisar áudios e, então, com as entradas de dados do usuário, indicar, em um gráfico, informações como “quem” está presente no material, áudio, entregue para a máquina. O algoritmo irá precisar de entradas, ou seja, um áudio de cada interlocutor para que possa armazenar e processar essas informações. O sistema será construído utilizando a linguagem Python e algumas de suas bibliotecas como *sklearn* que irá auxiliar na construção da rede neural já que possui diversas funções preparadas para diversos cenários e tipos de redes neurais, também serão utilizadas bibliotecas como *numpy* e *pandas* para a manipulação dos dados e *librosa* e *soundfile* para a extração dos atributos de cada arquivo de áudio, com essas últimas 4 bibliotecas citadas em mãos será possível desenvolver a parte mais pesada do algoritmo, que serão as funções de pré-processamento dos dados. Com as ferramentas em mãos agora precisamos dos dados e

para isso será utilizado o material de áudio da Open Speech and Language Resources, que pode ser encontrado no link: <http://www.openslr.org/12/>. A Open Speech and Language Resources disponibiliza uma grande variedade de arquivos de áudios gravados em estúdio e áudios gravados com interferência de outros ruídos. Para o desenvolvimento do projeto, usarei apenas os arquivos de áudios gravados em estúdio para facilitar o desenvolvimento e atestar a precisão do algoritmo, todos os arquivos de áudio utilizados possuem a extensão “*flac*”.

Além do objetivo principal, outro desafio é fazer com que o algoritmo faça o processamento dos arquivos de áudio em pouco tempo; para isso, será necessário implementar um tipo de processamento de dados, utilizando paralelismo, para obter a maior eficiência possível a fim de executar a predição em pouco tempo.

4 ANÁLISES E DISCUSSÕES

Durante o desenvolvimento, o algoritmo apresentou uma grande exigência por poder de processamento pesado, o que naturalmente pode se tornar um problema, então, o principal foco de início é fazer com que o algoritmo realize o processamento da forma mais eficiente possível e grave esses dados em uma base para que não seja necessário fazer esse processamento pesado em memória, toda vez que o algoritmo for executado.

Por isso, foi necessário realizar alterações no algoritmo para que utilizasse multiprocessamento (multithreading) da CPU a fim de aumentar a velocidade e também o poder de processamento. Mesmo assim, ainda demora muito tempo para que o pré-processamento dos dados seja finalizado e, então, a rede neural seja executada; em torno de 10 mins com aproximadamente 600 arquivos de áudio para treino, 1300 para teste e mais 600 para validação.

Vale ressaltar está sendo utilizado um Modelo Sequencial (Sequential Model) e Entropia Cruzada Categórica (Categorical crossentropy) durante o processamento de uma rede neural feedforward o que significa, que a rede neural não faz o ajuste automático do peso de cada neurônio, ou seja, é preciso de uma interação humana para realizar os ajustes finos de cada camada da rede neural como quantidade de neurônios e taxa de Dropout.

O modelo sequencial é apropriado para uma pilha plana de camadas, em que cada camada tem exatamente um tensor de entrada e um tensor de saída. Ou seja, para a rede neural realizar seus trabalhos, ela foi passada em uma lista contendo todos os arquivos de áudio e seus atributos; então, ao serem classificados e processados, a rede neural retorna uma lista com os dados processados e suas precisões.

Já a Categorical crossentropy, é uma função de perda utilizada para tarefas de classificação de múltiplas classes. Essas são tarefas em que um exemplo só pode pertencer a uma das muitas categorias possíveis e o modelo deve decidir qual delas. Isso é exatamente o que a rede neural faz em seu interior, de maneira que é o treinamento, teste e validação com 30 interlocutores diferentes.

O algoritmo irá possuir em seu interior duas classes e a rede neural, sendo as duas classes, para o pré-processamento dos dados, onde uma para realizar a leitura e categorização dos arquivos de áudio e outra para realizar a extração das informações de cada arquivo de áudio categorizado. E então, após a realização do pré-processamento e armazenagem dos dados em arquivos csv, a Rede Neural FeedForward entrará em ação para realizar o processamento dos dados salvos nos csv's e então emitirá uma imagem contendo um gráfico que irá mostrar a precisão de treinamento e de validação.

Foram feitas mais algumas melhorias para deixar o algoritmo mais eficiente ao realizar o pré-processamento com multithreading da CPU. Nesse contexto, serão feitas algumas melhorias no algoritmo de pré-processamento para criar um arquivo csv com os dados processados para cada caso, como treinamento, teste e validação a fim de agilizar a leitura dos dados e evitar um pré-processamento desnecessário dos mesmos dados, o que acaba por consumir excessivamente os recursos da máquina que o está executando;

Com as otimizações feitas, esses mesmos arquivos csv recebem os dados extraídos de cada áudio de cada caso em um objeto JSON, o que, para o algoritmo de pré-processamento acaba por agilizar o processo de processamento, uma vez que os dados já estão processados, categorizados e organizados pois ao ler o arquivo csv, os dados são convertidos para o objeto utilizado para a extração dos dados dos arquivos de áudio o que já os deixa em prontidão para a utilização dos mesmos pela rede neural.

A Rede Neural terá uma camada de entrada de 193 neurônios com a função de ativação da unidade linear rectificada ('relu') e sem valor definido para os bias. A camada seguinte terá uma entrada de 193 neurônios mas executará a função de Dropout com uma taxa de 0.1.

A camada Dropout define aleatoriamente unidades de entrada a 0 com uma frequência de taxa em cada etapa durante o tempo de treino, o que ajuda a evitar o excesso de equipamento. As entradas não ajustadas a 0 são aumentadas em $1/(1 - \text{taxa})$ de modo a que a soma sobre todas as entradas permaneça inalterada. (TENSORFLOW, 2022).

A próxima camada terá 128 neurônios de entrada e com função de ativação 'relu' e sem valor definido para os bias. Essa camada é seguida de outra com um Dropout de 0.25.

A camada seguinte terá 128 neurônios de entrada e com função de ativação 'relu' e sem valor definido para os bias. Essa camada é seguida de outra com um Dropout de 0.5.

A camada seguinte terá 128 neurônios de entrada e com função de ativação 'relu' e sem valor definido para os bias. Essa camada é seguida de outra com um Dropout de 0.5.

E por fim, a última camada, terá 30 neurônios de entrada com a função de ativação Softmax, que irá fazer a conversão do vetor de valores recebidos pelos neurônios para uma distribuição de probabilidade onde os elementos do vetor de saída estarão no intervalo (0, 1).

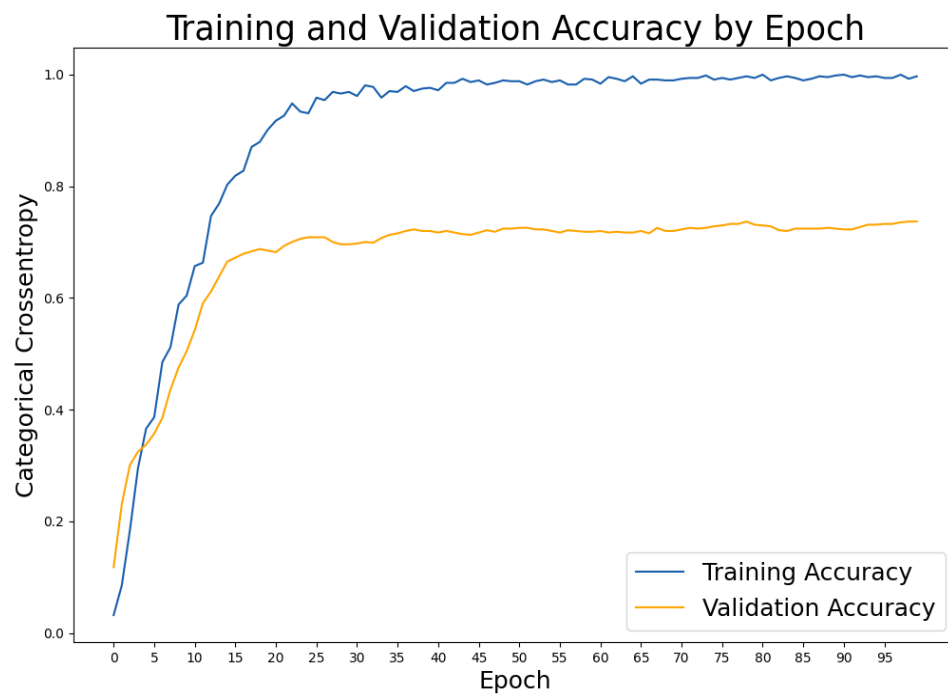
A ideia inicial seria criar um container docker com um banco de dados para aumentar a portabilidade e facilitar o acesso dos dados de qualquer lugar, simplesmente iniciando uma nova instância do banco via docker, porém após os ajustes de performance do algoritmo, o mesmo se mostrou totalmente capaz de realizar execuções rápidas e precisas utilizando arquivos csv. E como a implementação de um banco de dados acarretaria em um grande esforço o que possivelmente exigiria mais tempo, essa ideia foi abandonada, tendo em vista que o algoritmo já conseguiu atingir bons níveis de performance.

5 CONSIDERAÇÕES FINAIS

Foram realizadas diversas execuções do algoritmo em vários estágios do desenvolvimento, o que colocou em prova a performance e a eficiência do algoritmo em diversos cenários, dentre eles, a velocidade de processamento do algoritmo e a quantidade de arquivos de áudio processados por execução. Mesmo com aproximadamente 2000 arquivos para treinamento, 1300 para testes e 700 para validação, totalizando 4000 arquivos de áudios, o algoritmo mostrou-se bem eficiente, utilizando todos os núcleos do processador para realizar o pré-processamento dos arquivos de áudio em aproximadamente 10 minutos. Quanto ao processamento da rede neural, como atualmente ela possui poucas camadas, o seu processamento também foi bem rápido, chegando a processar os dados em torno de 1 minuto.

Com mais alguns ajustes finos, nas duas primeiras execuções, a rede neural obteve uma precisão de 73% na identificação dos interlocutores.

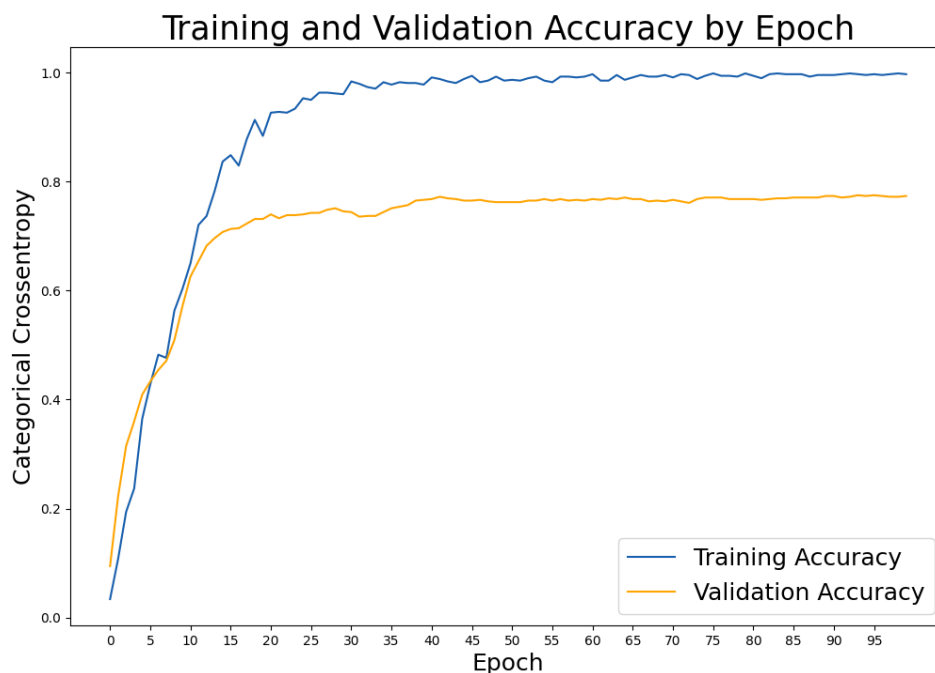
Figura 3: Gráfico apresentando a precisão de aproximadamente 73%.



Fonte: O autor, 2022.

Após 5 execuções da rede neural, foi capaz de alcançar uma taxa de precisão de 77% na identificação dos interlocutores.

Figura 4: Gráfico apresentando a precisão de aproximadamente 77%



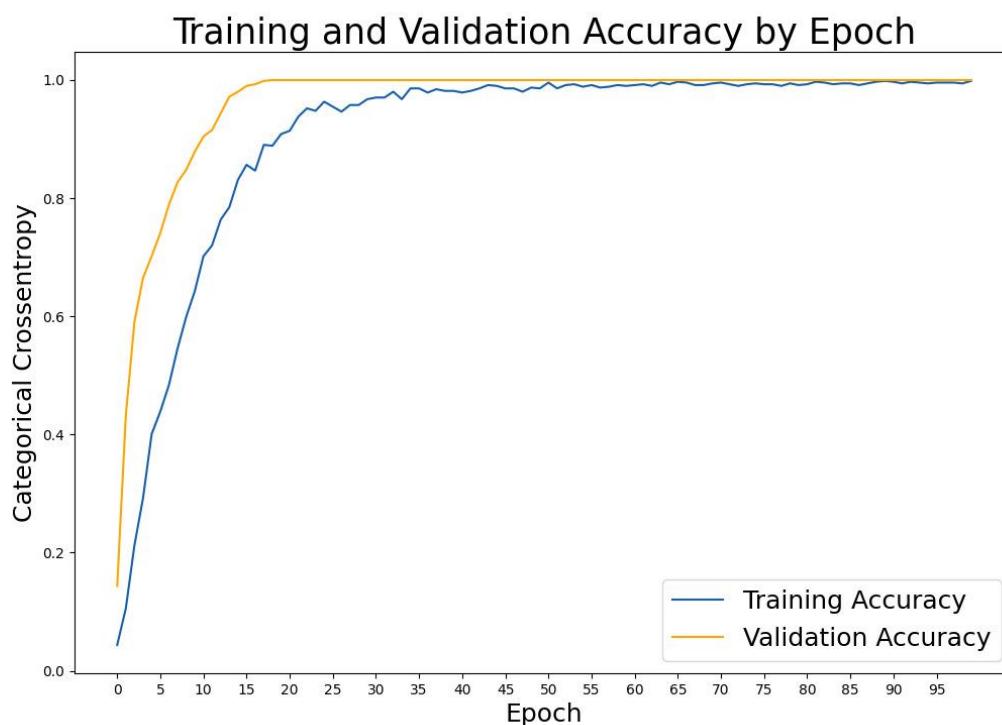
Fonte: O autor, 2022.

Pode-se observar que a diferença entre os dois gráficos é de aproximadamente 4%, o que significa que, com mais treinamento, a rede neural poderá ser capaz de alcançar uma maior precisão na identificação dos interlocutores.

Após realizar esses testes, podemos constatar que mesmo na figura 3 e 4 tivemos uma precisão excelente, por que os dados utilizados para treinamento eram ligeiramente diferentes dos dados utilizados para validação, ou seja, os 33% de diferença entre o treinamento (99% aproximadamente) e a validação (77%) da figura 4, apontam que 33% dos arquivos de áudio utilizados para validação possuem interlocutores diferentes dos utilizados durante o treinamento, ou seja, a rede neural reconheceu esses interlocutores por que ela não foi treinada para reconhecer eles, o que indica que o algoritmo acertou aproximadamente 99% dos interlocutores de validação que possuíam a mesma voz que foi utilizada para treinamento.

Contudo para atestar sua precisão foi necessário realizar um teste simples, mas com certeza muito efetivo, que foi utilizar exatamente os mesmos arquivos de validação dos treinamentos anteriores e para o treinamento e validação da rede neural, ou seja, utilizar exatamente os mesmos dados para treinamento e validação, e mesmo contendo 30 interlocutores diferentes, foi possível alcançar 99.86% de precisão na identificação dos interlocutores.

Figura 5: Gráfico apresentando a precisão de aproximadamente 99.86%



Fonte: O autor, 2022.

E para garantir ainda mais a precisão e também confirmar se ocorreu um fenômeno chamado de overfitting foram realizados diversos testes, inclusive adicionando mais camadas intermediárias contendo 128 neurônios com a função de ativação ‘relu’ e de Dropout com taxa de 0.3, 0.5, 0.6 e 0.9 na rede neural.

Um cenário de overfitting ocorre quando, nos dados de treino, o modelo tem um desempenho excelente, porém quando utilizamos os dados de teste o resultado é ruim. (Didática Tech, 2022).

Todos os resultados obtidos estão disponíveis no link a seguir: https://github.com/Bradoqguido/Neural_Network_To_Voice_Classification/tree/main/VoiceClassification/out.

Com todas essas informações podemos constatar que o algoritmo é sim efetivo e capaz de identificar as vozes de diversos interlocutores em arquivos de áudio e inclusive, mesmo que tenham áudios de interlocutores diferentes, a rede neural não irá se confundir facilmente, claro que não é sensato descartar a possibilidade de que a rede neural se confunda, pois pensar em um sistema que seja 100% à prova de falhas é praticamente impossível.

Além disso, como se trata de uma rede neural feedforward cada nova execução terá um resultado diferente, a menos que o usuário que está manipulando ela, altere as taxas de

dropout para influenciar o resultado final. Para evitar possíveis confusões e influências durante a manipulação do algoritmo é necessário evoluir a construção desta rede neural para algo mais robusto com feedforward e backpropagation o que resultará em um ajuste dos pesos de cada neurônio de forma automatizada pela própria rede neural a cada nova execução, evitando assim, erro humano e aprimorando ainda mais a rede neural e sua base de conhecimento o que por sua vez a tornará mais precisa e eficiente, tendo em vista que ela irá aprender a cada nova execução.

REFERÊNCIAS

APARECIDA, S. **Inteligência Artificial no Direito**. Disponível em: <https://jus.com.br/artigos/89283/inteligencia-artificial-no-direito>. Acesso em: 26 mai. 2022.

WIDROW, B. *et al.* **Adaptive “Adaline” neuron using chemical “memistors”**. **Number Technical Report 1553-2**. Stanford: Electron. Labs., 1960.

BARRETO, J. M. **Inteligência Artificial no Limiar do Século XXI**. 3ª ed. Santa Catarina: Duplic Prestação de Serviços, 2001.

BERETTA, L. **Inteligência artificial acelera processos no meio jurídico e estimula trabalho estratégico de profissionais do setor**. Disponível em: <https://www.ibm.com/blogs/ibm-comunica/inteligencia-artificial-acelera-processos-no-meio-juridico-e-estimula-trabalho-estrategico-de-profissionais-do-setor/>. Acesso em: 08 mai. 2022.

BRAGANÇA, F. BRAGANÇA, L. F. **Revolução 4.0 no Poder Judiciário: Levantamento do uso de inteligência artificial nos tribunais brasileiros**. Rio de Janeiro, v. 23, n. 46, p. 65-67, jul./out., 2019.

CARLOS, A. **Redes Neurais Artificiais são técnicas, adquirem conhecimento através da experiência**. Disponível em: <https://sites.icmc.usp.br/andre/research/neural/#:~:text=Redes%20Neurais%20Artificiais%20s%C3%A3o%20t%C3%A9cnicas,adquirem%20conhecimento%20atrav%C3%A9s%20da%20experi%C3%Aância>. Acesso em: 14 mai. 2022.

CARVALHO, T. **[LETRAS] FONÉTICA FORENSE: O PROFISSIONAL DE LETRAS COMO INVESTIGADOR CRIMINAL**. Disponível em: <https://www.literaturablog.com/letras-fonetica-forense-o-profissional-de-letras-como-investigador-criminal/>. Acesso em: 18 jun. 2022.

DE OLIVEIRA, R. F. **Inteligência Artificial**. Paraná: Editora e Distribuidora Educacional, 2018.

Didática Tech. **Underfitting e Overfitting**. Disponível em: <https://didatica.tech/underfitting-e-overfitting/>. Acesso em: 17 nov. 2022.

EDUARDA, M. A **Inteligência Artificial e Seu Impacto na Advocacia**. Disponível em: <https://ambitojuridico.com.br/cadernos/outros/a-interligencia-artificial-e-seu-impacto-na-advocacia/>. Acesso em: 26 mai. 2022.

FERNANDES, A. M. R. **Inteligência Artificial: Noções Gerais**. Florianópolis: Visual Books, 2003.

GIL, A. C. **Como elaborar projetos de pesquisa**. 6. ed. São Paulo: Atlas, 2017.

GULLY, A. *et al.* **How Voice Analysis Can Help Solve Crimes**, 2022. Disponível em: <https://kids.frontiersin.org/articles/10.3389/frym.2022.702664>. Acesso em: 15 mar. 2022.

JUDICIÁRIO. **Inteligência artificial está presente em metade dos tribunais brasileiros, aponta estudo inédito**. Disponível em: <https://www.stj.jus.br/sites/portalp/Paginas/Comunicacao/Noticias/09032021-Inteligencia-artificial-esta-presente-em-metade-dos-tribunais-brasileiros--aponta-estudo-inedito.aspx>. Acesso em: 08 mai. 2022.

MOREIRA, E. de S. **OS NEURÔNIOS, AS SINAPSES, O IMPULSO NERVOSO E OS MECANISMOS MORFO-FUNCIONAIS DE TRANSMISSÃO DOS SINAIS NEURAIS NO SISTEMA NERVOSO**. Disponível em: <http://editora.unifoa.edu.br/wp-content/uploads/2017/04/Volume-02.pdf>. Acesso em: 18 abr. 2022.

NILS, N. J. **Artificial intelligence: a new synthesis**. San Francisco: Morgan Kaufmann, 1998, p.1.

OZORIO, J. **Inteligência artificial bate 20 advogados em testes de revisão de contratos**. Disponível em: <https://www.conjur.com.br/2018-nov-21/inteligencia-artificial-bate-20-advogados-revisao-contratos>. Acesso em: 26 mai. 2022.

PÁDUA, E. M. M. **Metodologia da pesquisa: abordagem teórico-prática**. 13º Ed. Campinas: Papirus, 2007.

PEREIRA, M. H.; ROBERTO, M.; CAVALIERE, R. S. **Português V: volume 1, p. 11 e 23**. Rio de Janeiro: Fundação CECIERJ, 2015.

PERITO DE ÁUDIO, **Perícia de voz: Comparação fonética**, 2021. Disponível em: <https://peritodeaudio.com.br/index.php/pericia-de-voz>. Acesso em: 15 mar. 2022.

PRESSMAN, R. S; MAXIM, B. R. **Engenharia de Software: Uma abordagem profissional**. São Paulo: AMGH Editora Ltda., 2016.

POOLE, D. L.; MACKWORTH, A. K. **Artificial Intelligence: Foundations of Computational Agents**. United States of America: Cambridge University Press, 2010.

PRODANOV, C. C.; DE FREITAS, E. C. **Metodologia do trabalho científico: Métodos e Técnicas da Pesquisa e do Trabalho Acadêmico**. 2º Edição. Rio Grande do Sul: Feevale, 2013.

RAMALHO, A. M. R. **ESTRUTURA BIOLÓGICA DO SISTEMA NERVOSO** .

Disponível em:

https://cesad.ufs.br/ORBI/public/uploadCatalago/14445016022012Elementos_de_Anatomia_Humana_aula_12.pdf. Acesso em: 18 abr. 2022.

RUSSELL, S.; NORVIG, P. **Inteligência Artificial**. Rio de Janeiro: Elsevier, 2013.

PELTARION, **Categorical Crossentropy**. Disponível em:

<https://peltarion.com/knowledge-center/modeling-view/build-an-ai-model/loss-functions/categorical-crossentropy>. Acesso em: 10 ago. 2022.

SAS, **O QUE SÃO REDES NEURAIS?**. Disponível em:

https://www.sas.com/pt_br/insights/analytics/neural-networks.html#:~:text=Redes%20neurais%20s%C3%A3o%20sistemas%20de,tempo%20%E2%80%93%20aprender%20e%20melhorar%20continuamente. Acesso em: 6 abr. 2022.

SILVA, R. A. C. **Inteligência Artificial Aplicada a Ambientes de Engenharia de Software: Uma Visão Geral**. Minas Gerais, Universidade Federal de Viçosa, 2005.

SOMMERVILLE, I. **Engenharia de Software**. São Paulo: Pearson Prentice Hall, 2011.

TensorFlow, **tf.keras.layers.Dropout**. Disponível em:

https://www.tensorflow.org/api_docs/python/tf/keras/layers/Dropout. Acesso em 01 dez. 2022.

THIOLLENT, Michel. **Metodologia de Pesquisa-ação**. 14. ed. São Paulo: Cortez, 2018.