Capítulo 1

Fake News - Conceitos, métodos e aplicações de identificação e mitigação

Pablo de Andrades Lima, Érico Marcelo Hoff do Amaral, Alex Dias Camargo, Jean Lucas Cimirro e Gérson de Munhos Concilio

Abstract

This meta-article describes a short course designado for the 18th Regional School of Computer Networks. It is intended to address the topic fake news on the perspective of information security, engineering of digital lies, and legislation. It also aims to address some projects, applications and services to combat fake news. Finally, it presents computational methods available to mitigate this very contemporary problem, as it is extremely important to be understood by the academic and scientific community.

Resumo

Este meta-artigo descreve um minicurso elaborado para a 18ª Escola Regional de Redes de Computadores que pretende abordar o tema fake news sobre a óptica da segurança da informação, da engenharia das mentiras digitais, legislações, projetos, aplicações e serviços de combate as notícias falsas. Para finalizar, são abordados alguns métodos computacionais disponíveis para mitigação desse problema tão contemporâneo, pois é de suma importância a ser compreendido pela comunidade acadêmica e científica.

1.1. Introdução

O presente trabalho, foi desenvolvido pelo Grupo de Pesquisas acadêmicas do Pampa sobre disseminação de notícias falsas na *Internet* (Pampa sem *Fake*). Atualmente, faz parte do Programa Universidade *Hacker* - *UniHacker*, sendo composto por alunos do Curso de Engenharia de Computação da Universidade Federal do Pampa, Campus de Bagé e busca estudar, pesquisar e desenvolver possíveis soluções digitais que possam mitigar a disseminação de notícias falsas na *Internet*.

Neste sentido, após pesquisas exploratórias iniciais no ano de 2018, fica exposto que a liberdade e o contexto democrático da *Internet* acabou se tornando um cenário

propício para a propagação de informações, tornando este um meio comumente utilizado para o compartilhamento de falsas notícias, as ditas *fake news*. No Brasil, a propagação desse tipo de informação já ocorre no meio digital, sendo intensificado nos últimos anos durante o período eleitoral [Ruediger et al. 2018].

Após uma análise sistemática das principais redes sociais [Resende et al. 2018], foi estimado que 48% da população brasileira usa o *WhatsApp* (serviço de troca de mensagens via *smartphone* para compartilhar e discutir notícias. O GPOPAI - Grupo de Pesquisa em Políticas Públicas para Acesso à Informação [GPOPAI 2017], em 2017, através de um levantamento, alertou que cerca de 12 milhões de pessoas difundiram notícias falsas sobre política no Brasil.

Outro aspecto importante de citar, trata da influência do "filtro bolha"na difusão de *fake news* nas mídias sociais [Sastre et al. 2018], que tem sido também um dos grandes estimuladores de "bolhas digitais". O "filtro bolha"é um conceito usado para intitular algoritmos que direcionam o acesso ao conteúdo baseado no perfil e hábitos do usuário, dando uma impressão de eficiência na busca mas restringem a maneira de como a pesquisa é realizada, sendo este método muito utilizado pelo *website* de buscas *Google* e pela rede social *Facebook* [Pariser 2011]. Com isso, é possível sugerir que a influência do filtro bolha leva os usuários a bolhas digitais de informações, as quais tornam crenças acima do fato verdadeiro, evocando o que se chama de período "pós-verdade", diminuindo a realidade atual dos fatos objetivando sustentar ideologias e opiniões próprias [Poubel 2018]. Atualmente, existe uma grande gama de serviços, aplicações, projetos e pesquisas e legislações sobre o tema *fake news*. Este documento de minicurso visa navegar por esses tópicos relacionando-os as notícias falsas, levando, assim, conteúdo teórico e prático sobre o tema.

1.2. A engenharia das mentiras digitais

Nesta seção será efetuada uma introdução sobre segurança da informação e sistemas, bem como um histórico sobre o termo *fake news*. Do mesmo modo, serão tratados conceitos relacionados as notícias falsas como filtro bolha, bolha digital e pós-verdade.

1.3. Fake News

Uma das autoras mais citadas quanto a conceituação e especificação do termo *fake news* o classifica como complicado, e considera como um ecossistema de informações com muito mais do que notícias [Wardle 2017]. O termo falso não descreve a complexidade dos diferentes tipos de desinformação, decompondo em três elementos: as motivações de quem cria este conteúdo; as formas como este conteúdo está sendo disseminado; e finalmente, os diferentes tipos de conteúdo que estão sendo criados e compartilhados, os quais são descritos na Tabela 1.1. Para [QUESSADA and PISA 2018] a frase atribuída a Goebbels, Ministro da Propaganda de Adolf Hitler, “uma mentira propagada mil vezes torna-se verdade” é um exemplo de como as *fake news* atuam no período pós-verdade e a velocidade que se espalham nas redes sociais para embasar opiniões.

O uso de robôs em redes sociais é um dos escopos de pesquisa apontado como influenciador na disseminação de notícias falsas. [Ruediger et al. 2018] alertam para preocupação em entender, filtrar e denunciar a disseminação de informações falsas na *Inter-*

Tabela 1.1. Classificação das *fake news*

	Tipo	Descrição
1	Sátira ou paródia	Não quer necessariamente causar mal, mas pode enganar o leitor
2	Falsa Conexão	A chamada da notícia não condiz com conteúdo apresentado
3	Conteúdo Enganoso	Uso mentiroso de uma informação para difamar outro conteúdo ou pessoa
4	Falso Contexto	O conteúdo é verdadeiro, mas é compartilhado com contexto falso
5	Conteúdo Impostor	Quando usa o nome de uma pessoa ou marca, mas afirmações irreais
6	Conteúdo Manipulado	O conteúdo verdadeiro é alterado para enganar o público
7	Conteúdo Fabricado	Informações 100% falsas e construídas para causas mal e espalhar boatos

[Wardle 2017]

net com a utilização de ferramentas automatizadas (robôs sociais) que usufruem de perfis falsos, não reais, se passando por seres humanos e participando ativamente de debates políticos de grande repercussão; disseminando informações falsas, promovendo *hashtags* e massificando postagens automatizadas que comprometem o debate espontâneo através de *softwares* que geram este tipo de conteúdo artificialmente.

No estudo sobre "filtros bolha"de [Sastre et al. 2018], foram citadas as mudanças realizadas pelo *Facebook* com implantação do sistema de *crowdsourcing*, ou colaboração coletiva, que define as prioridades dos *feeds* de notícias nos perfis dos usuários, classificando o que irá aparecer ao usuário por uma maior familiaridade com os conteúdos mais acessados, com propósito de reduzir a difusão de *fake news* através de robôs. Porém, essa configuração gerou uma repercussão negativa com empresas que utilizam estratégias de divulgação por meio de mídias digitais. De acordo com Sérgio Dávila, editor-executivo do jornal a Folha de São Paulo, considerado o maior jornal do Brasil, as redes sociais tendem a criar "bolhas"e "condomínios de convicções"forçando as pessoas a se relacionar somente com outras que pensam como elas [Caulyt 2018].

A pesquisa de [Ferrara et al. 2016] mostra uma tendência promissora de evolução no combate às *fake news* utilizando o padrão automatizado de *Machine Learning* (ML) e inteligência humana para diferenciar robôs de pessoas. Não menos obstante, cabe citar o crescimento das *deepfakes*, uma técnica usada para substituir rostos originais em vídeos e utilizados para disseminar notícias falsas com uso de aplicativos para a troca de rosto, gerando grande quantidade de vídeos sinteticamente manipulados e distribuídos nas redes sociais, representando um grande desafio técnico para detecção e filtragem de tal conteúdo [de Moraes 2019].

O uso de perfis falsos para disseminação de notícias acaba sendo também um objeto de estudo de suma importância, pois esta técnica que é usada por empresas impulsionadoras de conteúdo ou pessoas comuns, muitas vezes tem como objetivo na propagação

de conteúdos falsos que são de seu interesse. Sabendo que a criação de perfis *fakes* é umas das formas mais utilizadas dentro das redes sociais para este fim específico, foi realizado um estudo baseado na criação de perfis falsos nas principais redes sociais, como, *Twitter e Facebook*, com a finalidade de identificar de que forma as redes tratam este tipo de prática, objetivando a mitigação de perfis que podem ser usados por robôs para disseminação de informações falsas e na criação em massa de contas em suas mídias.

Embora algumas redes sociais não possuam vínculo entre si, é notável a existência de um padrão de verificação de usuários durante a criação de perfis dentro das mesmas, onde é necessário efetuar a entrada com um *e-mail* para realizar a verificação inicial a qual já é um fator de validação. Com isso em vista, durante o procedimento de criação de contas *fake* foi utilizado o serviço chamado *Temp-Mail*[1], no qual foram gerados *emails* temporários que foram utilizados para realizar a ativação de tais perfis. Em ambas as redes sociais, em casos aleatórios, foi preciso de uma outra confirmação, por telefone, nestes casos foi necessário o uso de outra ferramenta, o aplicativo *2ndline - US Phone Number*, disponível para *Android* no *Play Store*, ao qual sua finalidade é disponibilizar um número de telefone celular norte-americano que recebe e encaminha mensagens SMS para a ativação de serviços *online*.

Nesse aspecto, foi notado que por questões publicitárias, ainda existem pequenas brechas para atuação de perfis falsos e até comércio deste tipo de contas conforme mostra (item 1) da Tabela 1.2 com *links* de acesso dos endereços. Há ainda no *Youtube* diversos vídeos e canais (item 2) que tratam da criação de perfis falsos para utilização em *marketing* digital, além de agências que prometem o impulsionamento digital através do uso de robôs (item 3).

Tabela 1.2. Lista de *links* de endereços

Item	Endereço	Descrição
1	http://www.fbpvastore.com/	Loja de Perfis
2	https://youtu.be/0eM8uFn1PAs	Canal de *marketing* digital
3	https://www.autland.com/	Agência de automação social

Desta forma, é possível perceber que o uso de perfis não verdadeiros ainda é um problema e, esta prática pode ser fortemente utilizada para quem tem por objetivo a disseminação de notícias falsas, pelo fato de que sua disseminação é muito rápida, principalmente quando esta notícia se é compartilhada em grupos de bolhas sociais, onde os assuntos são filtrados e os usuários recebem apenas determinados tipos de conteúdo. Sendo assim é importante que as pessoas tenham algum tipo educação digital para que possam confirmar em canais confiáveis de informação se determinado assunto é ou não verdadeiro e, se mesmo após essa reflexão ainda existir alguma dúvida, ir buscar informações a respeito para determinar se esta notícia é verídica.

1.4. Legislação

As tecnologias de informação e comunicação, assim como ambientes computacionais disponibilizaram nos últimos anos um conjunto de soluções para os usuários, como com-

[1] https://temp-mail.org

putadores de alto desempenho, dispositivos móveis, redes com altas taxas de velocidade e a *Internet*. Contudo, a utilização destes recursos nem sempre está direcionada a atividades lícitas, muitas vezes sendo estes mecanismos empregados em diferentes práticas ilegais, dentre elas se destacam a difusão de informações falsas e não verídicas. Nesse contexto, a presente seção deste minicurso relaciona a disseminação de noticias falsas com um dos mais importantes princípios da área de segurança da informação, a Legalidade, a qual descreve que o uso da Tecnologia de Informação e Comunicação deve seguir as leis vigentes do local ou país.

A seguir serão apresentadas, em relação as *fake news*, a interpretação do Código Penal Brasileiro sobre o tema, o Projeto de Lei 2360, o Código Eleitoral através da Lei 13.834 de 04 de junho de 2019, o Marco Civil da Internet (Lei nº 12.965, de 23 de abril de 2014) e a LGDP, Lei nº 13.709

O termo do inglês "*fake news*", ou seja, "notícias falsas"o qual se refere a qualquer notícia ou informação não verídica (falsa ou mentirosa) compartilhadas através de meios digitais como se verdade fosse. Esse tipo de mensagem é comumente disponibilizada por meio de redes sociais, aplicativos ou qualquer outro meio de comunicação na Internet. Neste contexto, o Código Penal brasileiro não previa tal disseminação como crime, visto que, Art. 5º, Inciso XXXIX, da Constituição Federal de 1988 proclama: "não há crime sem lei anterior que o defina, nem pena sem prévia cominação legal"(Princípio da Legalidade). Desta forma, o Código Penal e sua legislação não define como crime a divulgação de notícias falsas, tanto pela ausência de previsão de seu tipo normativo, assim como pela ausência de qualquer cominação de pena.

Contudo, o Projeto de Lei n° 2630, de 2020, "estabelece normas relativas à transparência de redes sociais e de serviços de mensagens privadas, sobretudo no tocante à responsabilidade dos provedores pelo combate à desinformação e pelo aumento da transparência na *Internet*, à transparência em relação a conteúdos patrocinados e à atuação do poder público, bem como estabelece sanções para o descumprimento da lei". Este projeto de lei, também denominada como Lei das *Fake News*, é um projeto de lei proposto pelo Senador Alessandro Vieira (CIDADANIA/SE) e definido como a Lei Brasileira de Liberdade, Responsabilidade e Transparência na *Internet*. Essa lei, já em seu primeiro artigo, "estabelece normas, diretrizes e mecanismos de transparência para provedores de redes sociais e de serviços de mensagens privada a fim de garantir segurança, ampla liberdade de expressão, comunicação e manifestação do pensamento."

O artigo 3° desta lei descreve algumas boas práticas, como medidas adequadas e proporcionais no combate ao comportamento inautêntico e na transparência sobre conteúdos pagos:

a. Liberdade de expressão e de imprensa;

b. Garantia dos direitos de personalidade, dignidade, honra e privacidade;

c. Respeito à formação de preferências políticas e de uma visão de mundo pessoal do usuário;

d. Compartilhamento da responsabilidade de preservação de uma esfera pública livre, plural, diversa e democrática;

e. Garantia da confiabilidade e da integridade de sistemas informacionais;

f. Promoção do acesso ao conhecimento de assuntos de interesse público;

g. Proteção dos consumidores; e

h. Transparência nas regras para anúncios e conteúdos patrocinados.

Além disso, o PL 2630 veda terminantemente: contas inautênticas, disseminadores artificiais não rotulados, redes de disseminação artificial que disseminem desinformação e conteúdos patrocinados não rotulados. A fim de atender Leis Eleitorais, este PL prevê a disponibilização a Justiça Eleitoral de dados relacionados a publicações politicas direcionadas a candidatos, partidos ou coligações, que tenham sido impulsionadas ou disseminadas, conforme descrito no art. 16 "Os provedores de redes sociais devem disponibilizar mecanismos para fornecer aos usuários as informações do histórico dos conteúdos impulsionados e publicitários"com os quais a conta teve contato nos últimos 6 (seis) meses."

Entre outros pontos importantes tem-se os art. 18 e 24 que tratam obre a responsabilidade de agentes políticos em mandatos eletivos e a proteção do direito de expressão e publicação de conteúdos por servidores públicos em suas contas privadas, de forma que não sejam perseguidos ou prejudicados por tais ações. Em relação as sanções, previstas neste Projeto de Lei, observa-se em seu Capítulo VI: art. 31. Sem prejuízo das demais sanções civis, criminais ou administrativas, os provedores de redes sociais e de serviços de mensagens privada ficam sujeitos a: I – advertência, com indicação de prazo para adoção de medidas corretivas; ou II – multa de até 10% (dez por cento) do faturamento do grupo econômico no Brasil no seu último exercício. §1º Na aplicação da sanção, a autoridade judicial observará a proporcionalidade, considerando a condição econômica do infrator, as consequências da infração na esfera coletiva e a reincidência. §2º Para os efeitos desta Lei, será considerado reincidente aquele que repetir no prazo de 6 (seis) meses condutas anteriormente sancionadas[2].

Em relação ao Código Eleitoral Brasileiro existe a Lei 13.834 de 04 de junho de 2019, a qual altera a Lei nº 4.737, de 15 de julho de 1965 (Código Eleitoral, para tipificar o crime de denunciação caluniosa com finalidade eleitoral). Esta lei pune com dois a oito anos de prisão quem divulgar notícias falsas com finalidade eleitoral. A lei havia sido sancionada originalmente em junho, mas um veto parcial tinha deixado de fora o dispositivo que criminaliza a disseminação de *Fake News* nas eleições.

Ao se tratar do Marco Civil da Internet torna-se claro o avanço desta Lei no trato jurídico das relações derivadas do uso da internet. Quando se discute a questão das *Fake News*, principalmente em períodos eleitorais (períodos normalmente rápidos e curtos), nota-se que a disseminação de noticias falsas pode culminar em alterações dos resultados de um pleito. Neste sentido é claro que o fator tempo é muito importante para se evitar problemas e ilícitos, neste contexto a utilização do *notice and take down* (baseia-se na notificação extra judicial para a remoção de noticias falsas no período de 24h, sendo o provedor solidário em responder por tal ilícito). Com a criação do Marco Civil da Internet

[2]https://www.politize.com.br/lei-das-fake-news/

passou-se a adotar o judicial *notice and take down*, ou seja, atualmente é necessária a notificação judicial para a remoção de determinado conteúdo do ar, o que amplia o tempo de resposta dos provedores, aumentando desta forma o fator de impacto do problema. Sendo assim, é notório que o Marco Civil da Internet é insuficiente em relação a propagação de notícias na Internet. Para ajustar estas brechas tem-se outros projetos como o Projeto de Lei n. 5.203 de 2016, que exige em sua redação, a indisponibilização por parte dos provedores de informações em um prazo de 48h após o recebimento de notificação[3].

Por fim, destaca-se a Lei Geral de Proteção de Dados (LGPD) a qual consiste na legislação brasileira que visa regulamentar a aquisição e o tratamento de dados pessoais no Brasil. Esta Lei por si só não é um instrumento específico para o combate a disseminação de *Fake News*, contudo permite a regulação do tipo e volume de dados que as empresas, detentoras de informações, poderão manipular e, com isso estar e conformidade com a lei específica para o combate a proliferação de notícias falsas, com base na segurança e salvaguarda de tais dados.

1.5. Aplicações e serviços de combate as notícias falsas

Nesta seção serão apresentados alguns projetos atuais, voltados para o combate a desinformação, os quais se baseiam na formação, educação midiática e conscientização, pontos estes considerados fundamentais para a mitigação do compartilhamento e notícias falsas e desinformação.

O primeiro serviço a ser abordado é o *Check*, uma ferramenta desenvolvida pela [Meedan 2020][4], que é uma organização sem fins lucrativos, foca na melhora da qualidade e equidade dos dados *online*, voltadas para redações, ONGs e instituições acadêmicas. É um espaço de trabalho *online* que permite aos usuários verificar fotos e texto *online* e criar conjuntos de dados. O *Check* ajuda jornalistas, organizações da sociedade civil, pesquisadores e investigadores de direitos humanos na vanguarda da coleta e verificação de informações. Durante o dia das eleições presidenciais de 2016 nos Estados Unidos, a *Check* apoiou o *Electionland*, um esforço nacional de cobertura de problemas de votação durante as eleições de 2016. A ferramenta utiliza uma estrutura de *workflow*, onde pessoas analisam e classificam dados, imagens e textos para uso na checagem de fatos. A empresa Meedan desenvolveu um conjunto de ferramenta para verificação de fatos interligada ao *WhatsApp, Facebook Messenger, WeChat* e outros aplicativos de troca de mensagens, ela permite que usuários encaminhem mensagens para um organização de verificação de fatos, atualmente em cinco países, Brasil, Africa do Sul, Índia, Quênia e Nigéria. Ao realizar uma solicitação os usuários receberão automaticamente os resultados dessa checagem de fatos, junto com algumas informações simples sobre por que a conclusão foi alcançada e um cartão visual que é projetado para ser compartilhável, a ferramenta pode receber solicitações em qualquer idioma, porém sua interface está atualmente disponível em inglês, espanhol, francês, português, árabe, russo e romeno. O principal diferencial do *Check* em comparativo a outras agências de checagem de fatos se dá pelo motivo de que uma informação que anteriormente foi verificada pelo seus colaboradores é armaze-

[3]https://jus.com.br/artigos/69900/a-insuficiencia-do-marco-civil-da-internet-em-relacao-as-fake-news-nas-eleicoes

[4]https://meedan.com/check

nada em um banco de dados e assim utilizando de *Machine Learning*, para fazer com que quando ocorra uma nova requisição por uma informação, primeiramente ela seja comparada com o conteúdo do banco de dados a fim de verificar se possui alguma reportagem ou informação similar ou igual a alguma que já foi verificada, fazendo com que a equipe de checagem não precise verificar novamente uma mesma informação fazendo com que os responsáveis por averiguar as informações tenham como foco em buscas ainda não classificadas.

O *Hoaxy*[5] é uma ferramenta *open source* desenvolvida pela Universidade de Indiana-IUNI, pelo CNetS e o *Observatory on Social Media*, com objetivo principal visualizar a disseminação de artigos online através do Twitter [Shao et al. 2016]. De acordo com autor a plataforma realiza coleta de dados, detecção e análise de desinformação online e direciona para verificação de fatos relacionados. Frisa que de modo geral, o compartilhamento de conteúdo de verificação de fatos geralmente fica atrás da desinformação entre 10 e 20 horas. Além disso, notícias falsas são dominadas por usuários muito ativos, enquanto a verificação de fatos é uma atividade mais básica. Com os riscos crescentes relacionados à enorme desinformação online, os observatórios de notícias sociais têm o potencial de ajudar pesquisadores, jornalistas e o público em geral a entender a dinâmica do compartilhamento de notícias reais e falsas.

Com algumas semelhanças, temos o *bot sentinel*[6], que trabalha com uso de aprendizado de máquina projetada para detectar robôs. O sistema pode classificar corretamente as contas com uma precisão de 95%, concentrando em comportamentos e atividades específicos considerados inadequados pelas regras do *Twitter*. Os pesquisadores analisaram contas que violavam repetidamente as regras do *Twitter* e treinaram o modelo para classificar contas semelhantes às que foram identificadas como "problemáticas". O *website* analisa centenas de *tweets* para classificar com precisão cada conta do *Twitter* e fornecer um relatório fácil de entender. Porém ele é incapaz de classificar se uma informação é verdadeira ou falsa pelo fato dele apenas comparar informações que são postadas por usuários com conta pública, a checagem é realizada apenas por máquina pois ele analisa apenas a disseminação das postagens e não seu conteúdo, pois o *bot sentinel* não possui um time humanizado de checagem.

Uma observação importante é que ideologia, afiliação política, crenças religiosas, localização geográfica ou frequência de *tweets* não são fatores para determinar a classificação de uma conta do *Twitter*. As contas são classificadas com base em uma pontuação de 0% a 100%; quanto maior a pontuação, maior a probabilidade da a conta se envolver em assédio direcionado, "trolagem tóxica"ou usar táticas enganosas projetadas para causar divisão e caos. Outro detalhe importante é análise de contas falsas que fazem parte de uma grande conspiração que tenta influenciar as políticas e/ou eleições ou disseminam campanhas de influência. Uma conta falsa que está ativamente tentando causar divisão e discórdia se comportará de maneira consistente com alguém que recebe uma pontuação alta do *bot sentinel*

Cabe citar ainda o NILC-USP – Detecção Automática de Notícias Falsas para o

[5]https://hoaxy.iuni.iu.edu
[6]https://botsentinel.com/

Português[7], desenvolvido por pesquisadores do Núcleo Interinstitucional de Linguística Computacional (NILC), da USP e da Universidade Federal de São Carlos (UFSCar), através de análise textual, com uso de inteligência artificial, para servir de apoio ao usuário na identificação de notícias falsas com uma precisão de 90% [Monteiro et al. 2018]. Ao receber um texto, o sistema aplica métodos para extrair atributos linguísticos desse texto e os utiliza em um modelo de aprendizado de máquina, que classifica a notícia como verdadeira ou falsa. O texto deve ter pelo menos 100 palavras. A partir destas palavras, ele irá verificar em dois modelos de detecção: Palavras do Texto e Classes Gramaticais. O primeiro modelo, é baseado em comparações sucessivas em um banco de 10395 palavras visando buscar palavras iguais as quais estão contidas no texto inserido pelo usuário, ao qual irá gerar uma pontuação para o texto informado, fazendo com que textos com uma maior ocorrência das palavras sejam mais plausíveis de veracidade. O segundo modelo irá calcular a porcentagem de classes gramaticais que está contida no texto a partir de uma biblioteca que está disponível na linguagem de programação Python(nlpnet [8]). Com as pontuações geradas pelos dois métodos, é aplicado um classificador que definirá se a notícia é verdadeira ou não. É importante ressaltar que os projetistas da ferramenta, não aconselham que checagem dos fatos seja feita apenas pela própria, pois a mesma usa apenas uma análise textual, sendo assim necessário uma humana para uma maior confiabilidade nos resultados.

Todas as ferramentas aqui descritas possuem algum método específico de atuação dentro de algum escopo, mas vale destacar [Monteiro et al. 2018] que também atua com uso de IA para classificação de notícias falsas.

1.6. Métodos computacionais disponíveis para mitigar

Nesta seção será feita uma explanação sobre Inteligência Artificial (IA) e Processamento de Linguagem Natural (do inglês, *Natural Language Processing*, NLP), assim como essas tecnologias podem ajudar computacionalmente no combate as *fake news*. Um exemplo prático implementado pode ser encontrado no repositório de códigos-fonte *GitHub* [9].

O termo Inteligência Artificial foi criado por [McCarthy et al. 1956] nos trabalhos do "*Dartmouth Summer Research Project on Artificial Intelligence*", os quais foi inserido como um novo campo de conhecimento associando linguagens e inteligência, raciocínio, aprendizagem e resolução de problemas. A comparação de humano e máquina a muito tempo é objeto de estudo, visto que [TURING 1950] já instigava reflexões como a questão: "Podem as máquinas pensar?", propondo o Jogo da Imitação. Em seus estudos, o autor esperava que máquinas e homens acabariam por competir em campos puramente intelectuais, porém encerrava com a afirmação: "Podemos avistar só um pequeno trecho do caminho à nossa frente, mas ali já vemos muito do que precisa ser feito"[TURING 1950].

Basicamente, um dos objetivos desta pesquisa é utilizar técnicas de Aprendizado de Máquina (AM), ou, em inglês, *Machine Learning* (ML). O tema trata de uma área da Inteligência Artificial que tem por objetivo a construção de sistemas computacionais capazes de adquirir conhecimento de forma automática, tomando suas decisões baseado

[7] https://nilc-fakenews.herokuapp.com/

[8] http://nilc.icmc.usp.br/nlpnet/

[9] https://github.com/alexcamargoweb/wrseg-2020/

em experiências acumuladas através de problemas anteriores com solução bem sucedida [Monard and Baranauskas 2003]. De acordo com [Murphy 2012], no AM, os algoritmos são capazes de realizar previsões de padrões em conjunto de dados pré-estabelecidos, podendo ser, essencialmente, classificados em dois tipos de aprendizado: supervisionado (preditivo) ou não-supervisionado (descritivo). Na abordagem aqui tratada foram utilizadas Redes Neurais Artificiais (RNAs), que são algoritmos (supervisionados) inspirados biologicamente no sistema nervoso de animais, úteis para processar grandes quantidades de dados de treinamento e objetivo de classificar novas instâncias [MARUMO 2018]. A arquitetura de uma RNA apresenta camadas de neurônios de processamento e conexões entre eles, onde durante o processo de treinamento são atribuídos pesos para cada conexão que servirão como determinantes de quais neurônios serão ativados na próxima camada, simulando as sinapses de um sistema nervoso animal [Bahdanau et al. 2014]. O processamento de cada neurônio é feito através de uma função de ativação que neste caso possui diversos tipos, onde a escolha desta função determina como devem ser as entradas da rede e como será o resultado de saída da mesma.

Com a evolução das RNAs, surgiu o termo *Deep Learning* (DL) ou Aprendizado Profundo, devido às muitas camadas por ela adotadas [Nallapati et al. 2016]. No contexto desta seção, esse tipo de aprendizagem, chamado Mineração de Texto, tende a identificar padrões de texto e analisá-los com uma técnica chamada *text summarization*. Posteriormente, os valores são classificados de acordo com as métricas pré-definidas e os pesos atribuídos a rede [MARUMO 2018].

Como parte de suma importância sobre esse tema, a mineração de texto pode ser compreendida pela técnica de Processamento de Linguagem Natural (conhecido amplamente por NLP). Essencialmente, trata da utilização de métodos e recursos computacionais para análise de dados linguísticos [Sakurai 2019]. Nesse contexto, uma implementação de NLP auxilia na resolução de ambiguidades estruturando numericamente o texto, separando palavras através num processo de "*tokenização*", processando-o com análises léxicas, sintáticas, semânticas e pragmáticas, para posterior compreensão do algoritmo de IA [Dale et al. 2000].

Em resumo, complementando e relacionando os conceitos aqui abordados, fica compreendido que dentro do conceito de IA, possuímos: o Aprendizado de Máquina (geralmente, supervisionado ou não supervisionado) como técnica de inferência; as Redes Neurais Artificiais como tipo de modelo preditivo e o *Deep Learning* como uma RNA com muitas camadas interligadas. Por fim, foi disponibilizado um *link* contendo uma implementação de uma NLP em *Python* via *Colab*, um ambiente em nuvem de alto desempenho mantido pelo *Google Research*.

1.7. Projetos sobre *Fake News*

Nesta seção, serão apresentados e listados na Tabela 1.3 alguns projetos que estão colaborando para o combate a desinformação atualmente. O Fato sem *Fake*-FsF, também da Universidade Federal do Pampa, é parte integrante de um projeto de pesquisa e extensão do Grupo *t3xto*. A proposta visa contribuir socialmente com conhecimentos sobre *fake news* e desinformação, para formar combatentes da "infodemia"que corrompe o bom senso democrático. O projeto é desenvolvido através de produção de *podcasts* apresen-

tados e produzidos pelo Prof. Marco Bonito e seus orientandos de iniciação científica: Gabriel Pujol, Luana Kasper e Emília Sosa, além do suporte do técnico de áudio Pedro Janczur. O projeto é financiado pela Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS).

Tabela 1.3. Projetos e ações de combate a notícias falsas

Nome	*Link* de acesso
[t3xto 2020]	https://anchor.fm/fato-sem-fake
[Paganotti et al. 2019]	https://vazafalsiane.com
[Sayad 2019]	https://educamidia.org.br

O projeto Vaza, Falsiane! é um curso *online* de iniciativa de três amigos, jornalistas e professores universitários que ao longo dos últimos dois anos estudam as *fake news*, investigando as melhores formas de produzir conteúdo sobre o assunto para um público amplo. Nesse período, foi incubado pela ONG Repórter Brasil e venceu um edital de financiamento do *Facebook*.

A EducaMídia, programa do Instituto Palavra Aberta com apoio do *Google.org* é uma entidade sem fins lucrativos que advoga a causa da plena liberdade de ideias, de pensamento e de opiniões. Com suas pesquisas, seminários e campanhas, promove a liberdade de expressão, a liberdade de imprensa e a livre circulação de informação como pilares fundamentais para o desenvolvimento de uma sociedade forte e democrática. O projeto Educamídia foi criado para capacitar professores e organizações de ensino, engajando a sociedade no processo de educação midiática dos jovens, desenvolvendo seus potenciais de comunicação nos diversos meios. Atualmente, atua no desenvolvimento de três competências: interpretação crítica das informações, produção ativa de conteúdos e participação responsável na sociedade. Possui atuação também na formação de professores e educadores, no apoio a formuladores de políticas públicas e na sensibilização para o tema. A plataforma centraliza conteúdos para formação e pesquisa, além de materiais e recursos para a sala de aula alinhados com a Base Nacional Comum Curricular (BNCC).

Outra ação de mitigação das notícias falsas, são as agências checagem de fato, *fact checking*, que dão credibilidade as notícias *online*. De acordo com [Fatos 2018], a checagem de fatos é um método jornalístico por meio do qual é possível certificar se a informação apurada foi obtida por meio de fontes confiáveis e, então, avaliar se é verdadeira ou falsa, se é sustentável ou não. Temos no Brasil hoje algumas agência de checagem de fatos como Aos Fatos[10] e Publica[11]. Alguns portais de notícias também tem oferecido esse tipo de serviço como o G1 que possui o portal Fato ou Fake[12] que atua de forma parecida as agencias de *fact checking*.

Encerrando esta seção e não menos importante, salientamos os esforços realizados pelo TSE-Tribunal Superior Eleitoral em combater as notícias falsas, que no meio político são comumente usadas. O portal da Justiça Eleitoral Brasileira, neste ano de 2020 com as eleições municipais, possui o projeto Fato ou Boato, criado em 2016 para ampliar

[10]https://www.aosfatos.org/

[11]https://apublica.org/

[12]https://g1.globo.com/fato-ou-fake/

o esclarecimento de informações relacionadas ao processo eleitoral, a página Fato ou Boato fomenta a circulação de conteúdos verídicos e estimula a verificação por meio da divulgação de notícias checadas, recomendações e conteúdos educativos.

Essa iniciativa integra o Programa de Enfrentamento a Desinformação nas Eleições 2020, que atualmente mobiliza mais de 50 instituições, entre partidos políticos e entidades públicas e privadas, para enfrentar os efeitos negativos provocados pela desinformação relacionada à democracia.

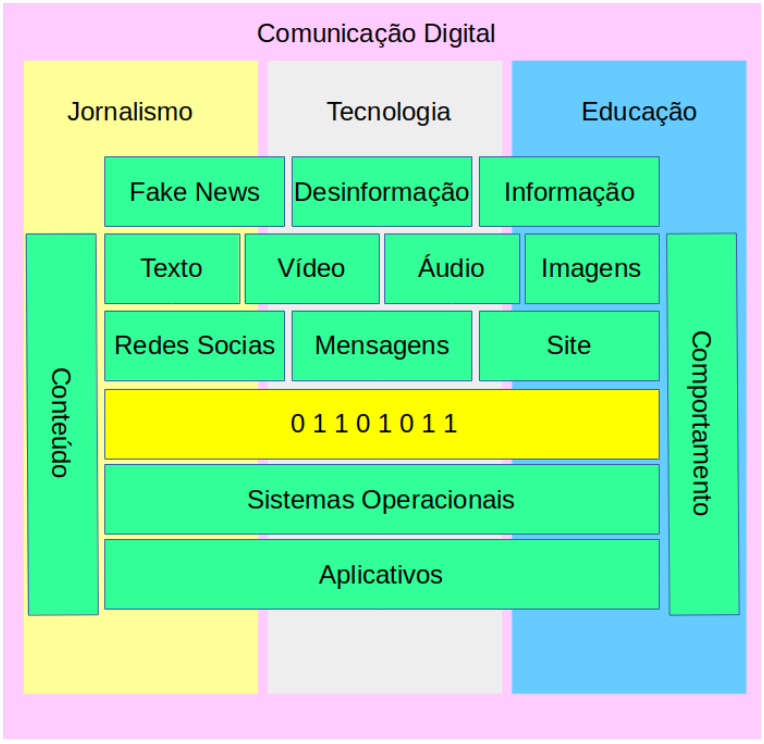
Tendo como maior objetivo o enfrentamento das *fake news*, nove das principais agências de checagem do Brasil integram essa força-tarefa em favor da circulação de conteúdos verificados, que efetivamente promovam debates e esclarecimentos fundamentais à tomada de decisão do eleitor.

Na centralidade da desinformação propagada em anos eleitorais, estão as notícias falsas sobre a urna eletrônica, que no portal dispões de diversas informações desmistificam dúvidas, fatos e boatos sobre o processo eleitoral eletrônico no Brasil. Outra ação importante desenvolvida pelo Tribunal, é a série de vídeos que apresenta de maneira simples e didática os passos que qualquer cidadão pode adotar para verificar conteúdos e se tornar um agente de combate à desinformação.

1.8. Considerações Finais

Uma das maneiras de poder visualizar o problema e suas possíveis soluções é a separação dele por "caixas", ilustrado na Figura 1.1 onde os possíveis métodos de solução atuariam especificadamente dentro de uma ou mais "caixas".

Figura 1.1. Diagrama sinóptico



O diagrama mostra ao fundo um problema de comunicação digital que no centro na verdade temos os dados em formato de *bits*, ou seja, a sua atuação e modo de operação é digital, tendo como áreas de atuação na sua possível solução, o jornalismo, educação e tecnologia. Podemos compreender também que se trata de um problema segmentado em *fake news*, desinformação e informação, onde esses dados transitariam em formatos de vídeos, texto, imagem ou áudio, por redes sociais, sites ou aplicativos de mensagens, sendo que estes poderiam ser tratados ou auditados através de uma análise de comportamento ou conteúdo com uso de aplicativos ou diretamente do sistema operacional.

Mesmo assim é possível verificar que boa parte das ferramentas tecnológicas que contribuem para redução da propagação de *fake news* através de robôs digitais, aproveitando do impulsionamento dos algoritmos de "filtro bolha", utiliza técnicas de *Deep Learning*.

O uso de perfis não verdadeiros ainda é um problema e esta prática vem sido fortemente usada para quem tem por objetivo a disseminação do notícias falsas. De acordo com [Ferrara et al. 2016], pesquisas mostram uma tendência promissora de evolução no combate às *fake news* utilizando o padrão automatizado de *Machine Learning* e inteligência humana para diferenciar robôs de pessoas.

Como métodos viáveis para a solução das *fake news*, partindo do pressuposto que a notícia falsa se propaga com maior proporção do que uma notícia verdadeira, é possível trabalhar também com a disseminação de fatos verídicos e, não somente com o combate à informações falsas. Mesmo assim, é possível verificar a dificuldade no tratamento de notícias falsas nos meios digitais atuais como redes sociais. Devido a complexidade do tema, legislações e algoritmo devem tomar todo cuidado no que tange a liberdade de expressão, sendo ainda a conscientização dos usuários a melhor forma de prevenção.

Referências

[Bahdanau et al. 2014] Bahdanau, D., Cho, K., and Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

[Caulyt 2018] Caulyt, F. (2018). Facebook perdeu importância para a folha. *diz editor. Deutsche Welle Brasil, versão online, Boon (Alemanha)*, 9.

[Dale et al. 2000] Dale, R., Moisl, H., and Somers, H. (2000). *Handbook of natural language processing*. CRC Press.

[de Moraes 2019] de Moraes, C. P. (2019). "deepfake" como ferramenta manipulação e disseminação de "fakenews" em formato de vídeo nas redes sociais.

[Fatos 2018] Fatos, A. (2018). O que é checagem de fatos—ou fact-checking. *Acesso em 5 de agosto de 2020*, 12.

[Ferrara et al. 2016] Ferrara, E., Varol, O., Davis, C., Menczer, F., and Flammini, A. (2016). The rise of social bots. *Communications of the ACM*, 59(7):96–104.

[GPOPAI 2017] GPOPAI (2017). Públicas para o acesso à informação. Technical report, da EACH/USP–GPOPAI. Subsídio público e acesso ao conhecimento. 2017 ....

[MARUMO 2018] MARUMO, F. S. (2018). Deep learning para classificação de fake news por sumarização de texto.

[McCarthy et al. 1956] McCarthy, J., Minsky, M., and Rochester, N. (1956). The dartmouth summer research project on artificial intelligence. *Artificial intelligence: past, present, and future*.

[Meedan 2020] Meedan (2020). Check.

[Monard and Baranauskas 2003] Monard, M. C. and Baranauskas, J. A. (2003). Conceitos sobre aprendizado de máquina. *Sistemas inteligentes-Fundamentos e aplicações*, 1(1):32.

[Monteiro et al. 2018] Monteiro, R. A., Santos, R. L. S., Pardo, T. A. S., de Almeida, T. A., Ruiz, E. E. S., and Vale, O. A. (2018). Contributions to the study of fake news in portuguese: New corpus and automatic detection results. In Villavicencio, A., Moreira, V., Abad, A., Caseli, H., Gamallo, P., Ramisch, C., Gonçalo Oliveira, H., and Paetzold, G. H., editors, *Computational Processing of the Portuguese Language*, pages 324–334, Cham. Springer International Publishing.

[Murphy 2012] Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press.

[Nallapati et al. 2016] Nallapati, R., Zhou, B., Gulcehre, C., Xiang, B., et al. (2016). Abstractive text summarization using sequence-to-sequence rnns and beyond. *arXiv preprint arXiv:1602.06023*.

[Paganotti et al. 2019] Paganotti, I., Sakamoto, L. M., and Ratier, R. P. (2019). "mais fake e menos news": resposta educativa às notícias falsas nas eleições de 2018. *Liberdade de Expressão Questões da atualidade*, page 52.

[Pariser 2011] Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. Penguin UK.

[Poubel 2018] Poubel, M. (2018). Fake news e pós-verdade. *Infoescola. Sociedade. s/d. Disponível em< https://www. infoescola. com/sociedade/fake-news*, 15.

[QUESSADA and PISA 2018] QUESSADA, M. and PISA, L. F. (2018). Fake news versus mil: a difícil tarefa de desmentir goebbels.

[Resende et al. 2018] Resende, G., Messias, J., Silva, M., Almeida, J., Vasconcelos, M., and Benevenuto, F. (2018). A system for monitoring public political groups in whatsapp. In *Proceedings of the 24th Brazilian Symposium on Multimedia and the Web*, pages 387–390.

[Ruediger et al. 2018] Ruediger, M. A., Grassi, A., and Guedes, A. L. (2018). Robôs, redes sociais e política no brasil: análise de interferências de perfis automatizados de 2014.

[Sakurai 2019] Sakurai, G. Y. (2019). Processamento de linguagem natural-detecção de fake news.

[Sastre et al. 2018] Sastre, A., de Oliveira, C. S. P., and Belda, F. R. (2018). A influência do "filtro bolha" na difusão de fake news nas mídias sociais: reflexões sobre as mudanças nos algoritmos do facebook. *Revista GEMInIS*, 9(1):4–17.

[Sayad 2019] Sayad, A. L. V. (2019). Educação midiática e pensamento crítico: antídotos contra a "desinformação". *Liberdade de Expressão Questões da atualidade*, page 9.

[Shao et al. 2016] Shao, C., Ciampaglia, G. L., Flammini, A., and Menczer, F. (2016). Hoaxy: A platform for tracking online misinformation. In *Proceedings of the 25th International Conference Companion on World Wide Web*, WWW '16 Companion, page 745–750, Republic and Canton of Geneva, CHE. International World Wide Web Conferences Steering Committee.

[t3xto 2020] t3xto, G. (2020). Pampa sem fake. Access date: 04 nov. 202020.

[TURING 1950] TURING, I. B. A. (1950). Computing machinery and intelligence-am turing. *Mind*, 59(236):433.

[Wardle 2017] Wardle, C. (2017). Fake news. it's complicated. *First Draft*, 16.