

# Enquadramento Legal ao Treino de Modelos de Linguagem Computacional - O Caso Português

Rúben Almeida - up201704618@fe.up.pt

Direito e Privacidade - FCUP em 12 Dezembro de 2022

## Abstract

A procura por padrões legislativos mais assertivos na área dos sistemas de informação criou uma dissonância entre o mundo jurídico e o mundo da ciência de computação. Uma barreira, desde logo técnica, pois os conhecimentos de ambos são disjuntos, mas também comunicacionais dado a linguagem técnica dos engenheiros informática divergir muito da formalidade do direito. O presente documento pretende endereçar essa tensão através da produção de uma análise da legislação que se aplica em Portugal ao trabalhar em Processamento de Linguagem Natural escrito numa linguagem de engenheiro endereçada a engenheiros.

## 1 Introdução

### 1.1 O Processamento de Linguagem Natural

A linguagem computacional, do inglês *Natural Language Processing (NLP)*, é o ramo da ciência de computadores que se debruça em dotar os computadores de capacidades de entendimento e produção de conteúdos em linguagem humana. Uma área de conhecimento com cinquenta anos de investigação, que está atualmente alicerçada sobre técnicas de redes neuronais.

### 1.2 Os Sucessos do NLP

Os sucessos mais conhecidos comercialmente do NLP são:

- Os Assistentes Pessoais: Siri/Alexa/Cortana
- O Google Tradutor
- Os Corretores Ortográficos do Microsoft Word

### 1.3 A Aprendizagem Computacional

Ao contrário da ideia que os órgãos de comunicação social construíram junto da sociedade, a inteligência artificial, do inglês *Machine Learning* não é um conceito novo.

A primeira referência ao termo acontece pelas mãos do cientista de computação, John McCarthy, em 1955, ao cunhar o termo como "a ciência e engenharia de fazer as coisas inteligentes" [1].

A confusão surge porque esta área das ciências computacionais não produzia até 2011 resultados brilhantes. Até essa data, os fracos desempenhos afastavam o interesse das grandes empresas, e por conseguinte os utilizadores. Tudo mudou na viragem da década, quando os computadores tornaram-se avançados o suficiente para suportar redes neuronais artificiais muito profundas.

### 1.4 As Redes Neuronais Artificiais Profundas

As redes neuronais artificiais profundas (RN) são uma das muitas técnicas usadas pela informática para con-

seguirem dotar os computadores de capacidades de aprendizagem. É de longe a técnica que melhores resultados produz. Essa qualidade de desempenho é feita à custa de muito poder computacional.

## 1.5 O Processo de Treino de RNs

O elemento chave para compreender os resultados espantosos das RNs requer entender o seu processo de treino.

De uma forma simplificada, considerando um sistema que distingue tweets ofensivos de tweets respeitadores, treinar uma RN capaz de fazer esta distinção é na realidade o treino de uma reta matemática que divide o universo em dois: os valores superiores à reta são de um tipo (ex: este Tweet é ofensivo) e os valores inferiores são de outro tipo (ex: este tweet é respeitador).

Este processo de aprendizagem automatizado da equação da reta de fronteira é feito sem intervenção humana, bastando o computador processar um grande volume de dados, um *Dataset*, para determinar qual as configurações que melhor separam os dois universos.

## 1.6 Datasets: Uma Necessidade

Como referido acima, os milagres da tecnologia (1.2) só são possíveis à custa de muitos dados. É nesta necessidade que a discussão moral-o-jurídica se inicia.

Para treinar um sistema de deteção de linguagem de ódio é necessário agrupar e disponibilizar ao computador uma coleção de textos de ódio e textos respeitadores. De forma análoga, para uma tarefa de aprendizagem capaz de distinguir cães de gatos, é necessário agrupar e disponibilizar um dataset composto por fotos de cães e de gatos ao computador.

## 1.7 Organização do Documento

O presente documento irá, feita a introdução técnica, organizar-se da seguinte forma:

- Secção 2 - Identificação dos Direitos Conflituantes ao Trabalhar em NLP

- Secção 3 - Apresentação de um Caso Prático: A Cambridge Analytica
- Secção 4 - A Proposta de Regulamento Europeu para IA
- Secção 5 - Conclusões

Sempre que existir jurisprudência pertinente para sustentar os elementos expostos, ela será apresentada inserida no decorrer do texto.

# 2 Direitos Conflituantes

## 2.1 Direitos de Utilização Informática

Os investigadores podem estar seguros dos direitos legais que a república portuguesa consagra aos programadores. A informática tem um destaque diferenciada na legislação portuguesa, começando logo pelo art.35º, alínea 1[2] da constituição(CRP) que consagra o direito à **Utilização Informática**.

### 2.1.1 Reserva ao Tratamento de Dados Sensíveis

O direito à utilização informática estipula, no entanto, na sua alínea 3[2], alguns limites ao que é possível fazer em informática e tratar em termos de dados. Os investigadores devem, desde logo, compreender que há um conjunto de temáticas de sensibilidade particular que devem, salvo em caso muito particular, abertamente consentido e de necessidade imperiosa, abster-se de trabalhar com:

- Convicções Políticas e Filosóficas
- Fé e Religião
- Vida Privada
- Questões Étnicas

Ou seja, os investigadores devem abster-se de criar recursos que incidam sobre estas questões, tais como modelos de previsão, datasets, ferramentas de análise ou extração de informação.

## 2.2 Direito De Autor

No âmbito do direito civil, o ramo dos direitos inatos de criação possui uma posição de importância. Esse facto é observável na existência de um documento específico regulador desta temática, o Código do Direito de Autor e dos Direitos Conexos.

### 2.2.1 Direito Inato À Criação

A legislação de "Direitos de Autor" não consagra somente direitos ao objeto da produção criativa. A CRP estabelece, desde logo, o direito inato ao Homem de criar obra[3]. Ou seja, no caso da investigação em NLP, o investigador sabe que pode contar com o direito a produzir um modelo, um software, um dataset ou outro conteúdo informático, independentemente de quem se trata, da literacia ou da filiação académica

### 2.2.2 Direitos Consagrados

Quando uma obra é abrangida pelo código dos direitos de autor passa a usufruir de um estatuto legal privilegiado que impede que um agente possa de livre vontade, sem consentimento explícito do autor, apropriar-se, reproduzir ou manipular a obra criada.

### 2.2.3 Conteúdos Cobertos

De acordo com Teresa Nobre[4], o objeto do direito de autor são somente **obras**. A lei compreende como uma obra uma "exteriorização de criação **intelectual**". Podendo as obras dividirem-se em duas categorias:

- **Obras Artísticas:**

- Pinturas, Desenhos, Vídeos
- Fotografia **artística**

- **Obras Literárias:**

- Jornais e Revistas, Livros
- Artigos Técnicos e Científicos
- Programas de Computadores, Código Fonte de Software
- Compilações Anotadas (Datasets)

### 2.2.4 Conteúdos Isentos

Como é possível verificar muitos das habituais fontes de dados utilizadas em NLP estão cobertas pelos direitos de autor no direito português. Os investigadores deverão estar desta forma atentos ao licenciamento associado aos datasets, se estes requerem de consentimento dos autores. Já corpus textuais como posts em blogs e redes sociais muito habituais no treino de modelos, não requerem, em geral, preocupações acrescidas por parte dos investigadores.

### 2.2.5 Licenças de Livre Acesso: A Creative Commons

Existe também casos em que o autor prescinde parcialmente dos seus direitos de autor no conteúdo que publica. Tipicamente no mundo digital isso acontece através da subscrição de convenções internacionais como a **Creative Commons**[5]. O caso mais conhecido de uma plataforma sem fins lucrativos que assina a política Creative Commons é a Wikipédia, desta forma é seguro investigadores utilizarem os conteúdos da Wikipédia nas suas atividades.

### 2.2.6 Dados Produzidos pela Máquina

À medida que a Inteligência Artificial adquire capacidades, esta assume relevância cada vez maior no dia a dia das sociedades. Face a este acréscimo de responsabilidades, duas questões se levantam: quem tem a responsabilidade civil no caso de falha destes agentes inteligentes artificiais? As obras produzidas por estes agentes possuem direitos de autor?

No caso do NLP, podemos mencionar as capacidades de geração de texto de todos os grandes modelos de linguagem, lembrando o GPT-3 e a sua capacidade de gerar um ensaio académico sobre ele próprio[6]. Coloca-se a questão a quem pertencem os direitos de autor deste documento?

Sobre estas questões, Mafalda Barbosa [7], aponta trabalho feito pela Juristische Fakultät da Universität Basel e conclui que não é provável que o direito venha a consagrar a figura jurídica de "E-Persons" aos agentes

inteligentes. Mantendo-se a questão bastante confusa e a carecer de análise profunda.

## 2.3 Direito à Palavra

Salvo casos muito específicos, a lei restringe a **captação e divulgação** da palavra de um cidadão através do artigo 79.<sup>o</sup> do Código Civil(CC)[8]. No mundo do direito, a palavra abrange quer a voz, quer a escrita.

### 2.3.1 Semelhanças com o Direito à Imagem

Na realidade, o artigo 79.<sup>o</sup> não incide somente sobre a palavra, mas também sobre a imagem dos cidadãos. É explícito no referido artigo que a palavra possui um enquadramento legal análogo à imagem[8]

O corolário da analogia é que o desenvolvimento para NLP ou para visão de computadores tem um regime jurídico semelhante para este artigo.

### 2.3.2 Diferença com Direitos de Autor

O direito à palavra diferencia-se dos direitos de autor, tendo a palavra uma cobertura legal mais larga além das obras intelectuais. Desta forma, as exceções(2.2.4)denotadas aos direitos de autor, como os posts de redes sociais, estão agora, protegidas, pelo direito à palavra.

### 2.3.3 Justificações da Violação

A legislação[8] estabelece que o direito à palavra pode ser violado sem censura nas seguintes condições:

- Palavra de Cargos Importantes Socialmente
- Palavra captada em espaço público
- Factos de interesse público
- **Finalidades Científicas, Didáticas e Culturais**

### 2.3.4 "A Finalidade Científica"

A legislação[8] isenta a atividade científica da necessidade de consentimento para o uso da palavra. Contudo isto levanta algumas questões, como:

- O que é que o Direito entende como Ciência?
- Existe Ciência Privada / Uma empresa com fins lucrativos pode produzir ciência?

As respostas para estas duas questões podem ser encontradas no decreto-lei n.º63/2019, de 16 de maio, onde se estabeleceu um regime jurídico específico para as instituições que se dedicam à investigação científica e desenvolvimento. Neste documento define-se ciência como "atividades criadores de conhecimento" [9].

O documento estipula que as instituições produtoras de ciência podem ser públicas ou privadas, não mencionando explicitamente se uma empresa de serviços informáticos, por exemplo, pode acumular, simultaneamente, atividades de ciência.

Desta forma, a sugestão é que exista uma clara definição que uma determinada instituição ou pessoa seja produtora de ciência de forma a reduzir as potenciais ambiguidades jurídicas nesta matéria. Esta parece ser a prática comum seguida pela Google ou Microsoft e as suas subsidiárias de NLP, a Deep Mind e OpenAI, respetivamente.

Por último, a alínea três do artigo 79<sup>o</sup>[8] que rege o direito à palavra estabelece uma **regra inviolável**: em caso algum poderá existir "prejuízo da honra, reputação ou simples decoro da pessoa" a quem se confere o direito à palavra.

### 2.3.5 Um Conflito Grave com NLP

Se as exceções enunciadas no capítulo acima(2.3.3) não estivessem legalmente consagradas seria praticamente impossível desenvolver em NLP de forma legal.

A exceção consagrada para a investigação científica é preponderante para estabelecer um enquadramento legal à atividade.

Por último, de notar a necessidade de requisitar sempre o consentimento legal aos utilizadores em

serviços de transcrição de voz, como, por exemplo, a Siri, pois existe a captura da voz do utilizador, o que por si só constitui um ilícito ao abrigo do artigo 70.º do CC.

## 2.4 Direito ao Carácter

O direito ao carácter estabelece os limites legais ao ato de *profiling*. Em particular, o Regulamento Geral de Proteção de Dados(RGPD) estabelece no seu artigo 22.º[10] que não são permitidas decisões totalmente automatizadas, nem a criação de perfis. Uma atividade muito usual pré-RGPD, onde perfis eram criados através do processamento de dados produzidos pelo agente dentro de plataformas digitais.

Desta forma, o investigador em NLP deverá estar sempre ciente que o produto do seu trabalho nunca poderá produzir uma decisão final sobre qualquer assunto, requerendo algures na cadeia de decisão a intervenção de um humano, para estar em harmonia com a legislação em vigor.

## 2.5 Direito à Proteção de Dados

A última grande produção legislativa na área das tecnologias de informação é o, sobejamente falado, Regulamento Geral de Proteção de Dados de 2016[11]. Este documento aplicado a nível europeu introduziu um conjunto largo de direitos aos cidadãos para fomentar maior transparência e responsabilidade no tratamento de dados pessoais.

Este regulamento é ainda complementado pela **Lei Nacional de Proteção de Dados**, Lei Nº58/2019 de 8 de agosto. Aplicando-se ambas em simultâneo e de forma complementar sobre a temática da proteção de dados

### 2.5.1 Conceito de Dados Pessoais

A legislação mencionada atua somente sobre Dados Pessoais. O diploma define dados pessoais como, "todos os dados que permitam identificar um cidadão, o titular dos dados"[11]. Sendo dado como exemplos,

nome, n.º Cartão de Cidadão, dados de localização, entre outros.

As sociedades comerciais, fruto de não serem cidadãs, não tem direitos reconhecidos no RGPD, sendo o tratamento de dados referentes às mesmas mais permissivo de um ponto de vista jurídico.

Ainda de referir a jurisprudência alemã que considerou o endereço de IP um dado pessoal, ou seja, um dado capaz de identificar um cidadão[12].

### 2.5.2 Conceito de Tratamento de Dados Pessoais

O diploma é muito abrangente no que considera tratamento de dados e engloba todas as tarefas tipicamente utilizadas por investigadores em NLP, desde a recolha ao armazenamento, passando pela filtragem e divulgação.

### 2.5.3 Direitos Consagrados

Com o RGPD, um cidadão europeu passa a poder usufruir de um conjunto de direitos na sua relação com a entidade tratadora de dados. Listam-se alguns: Transparência no Uso Dados, Necessidade de Consentimento Prévio, Direito ao Acesso aos Dados, Direito ao Esquecimento, Proibição de Tratamento de Dados Sensíveis ...

### 2.5.4 Minimização dos Dados

O legislador europeu estabeleceu um conceito novo neste regulamento intitulado de **minimização dos dados**. Segundo este princípio, os investigadores devem abster-se de recolher dados fora do âmbito estritamente necessário para cumprir os seus objetivos.

### 2.5.5 Localização Geográfica dos Dados

Uma questão que se coloca é se o RGPD se aplica quando o servidor onde os dados estão armazenados em jurisdições se encontra fora da União Europeia. O diploma no seu artigo 4.º[11] é muito claro que sim. A regra é simples, se uma entidade de qualquer nacional-

idade opera na UE, ela tem de cumprir as regras da UE.

### 2.5.6 Cobertura de Dados Automatizados-O Caso Google Spain

Durante o caso Google Spain[13] foi discutido se o carácter automático da indexação realizada pela Google exclui a empresa de responsabilidade pelos conteúdos tratados. O tribunal decidiu em favor do queixoso, tendo rejeitado a argumentação da Google invocando a dificuldade de controlar as ações automáticas.

Desta forma, os investigadores de NLP devem estar cientes que o comportamento dos seus modelos deve ser passível de algum nível de supervisão, pois atividades ilícitas podem ser imputadas aos seus criadores. A aleatoriedade e automatização dos modelos de Deep Learning não é por si mesmo um motivo de exclusão de responsabilidade civil.

### 2.5.7 Tratamento de Dados Para Investigação Científica

O artigo 89.º do RGPD[11] prevê que a investigação científica pode utilizar dados pessoais sem consentimento expresso desde que siga padrões éticos e cumpra o princípio da minimização de dados(2.5.4).

Ou seja, os investigadores têm aqui um critério relaxado e espaço para desenvolverem os seus trabalhos.

## 2.6 Comentário à Escassez de Jurisprudência

Uma das maiores dificuldades encontradas no decurso deste trabalho foi a escassez de jurisprudência capaz de sustentar algumas das considerações extraídas da legislação.

A pesquisa nas plataformas que indexam jurisprudências, tais como a dgsi[14], ou a plataforma europeia EURO-Lex[15], retornam poucos ou nenhuns resultados quando se pesquisa por termos como **Machine Learning, Processamento de Linguagem Natural, Redes Neurais**.

Este facto é indicativo que ainda não se tornou prática comum aos tribunais portugueses e europeus o pronunciamento sobre este tipo de temáticas. Contudo, é expectável que em breve essa realidade se altere, aumentando as fontes de conhecimento sobre como operar dentro da legalidade neste setor.

## 3 Caso Prático: O Estranho Caso da Cambridge Analytica

O caso da Cambridge Analytica(CA)[16] marcou 2018, fundada em 2013 em Londres no Reino Unido, esta empresa especializou-se na exploração dos dados do Facebook, tais como, likes, posts, fotos e localização.O objetivo era a produção de perfis políticos(2.4) de milhões de pessoas em todo o mundo com o intuito de vender um serviço de propaganda política adaptada às preferências de cada cidadão, a campanhas eleitorais, um pouco por todo o mundo.

### 3.1 A Crítica Apontada

De acordo com muitos agentes sociais, este nível de personalização é profundamente nefasto para a manutenção da democracia. São apontadas a criação de bolhas sociais, a falta de contraditório e do debate de ideias, como causas para polarizações políticas extremas.

Além disso, a propaganda personalizada que era debitada às vítimas era tipicamente falsa, imprecisa e extremamente sensacionalista.

### 3.2 Papel dos Envolvidos

O caso da CA que operava como entidade tratante dos dados, envolveu ainda o Facebook. Foi esta rede social a fonte de dados utilizados pela CA, tendo obtido ganhos financeiros pela venda dos mesmos.

Dentro dos clientes da CA contabilizam-se Ted Cruz[17], Donald Trump[18] ou o movimento pela saída do Reino Unido da UE, o UKIP[19].

Para treino dos modelos preditivos e alojamento do enorme volume de dados recolhido, a CA utilizava

serviços de alojamento dos serviços digitais da Amazon, a AWS.

### 3.3 O Papel do NLP na CA

A CA orientou o seu modelo de negócio em torno da capacidade de extrair informações acerca das suas vítimas utilizando publicações textuais, tipicamente posts de Facebook. Técnicas automatizadas de NLP como análise de sentimento ou de extração de tópicos e de entidades eram nucleares para o sucesso desse modelo de negócio.

### 3.4 Direitos Violados

Muitos foram os direitos que se suspeita terem sido violados. Quer pela CA, quer pelo Facebook.

Pelo Facebook:

- Direito à Proteção de Dados-Várias cláusulas do RGPD foram violadas, a começar pela ausência de consentimento para que os dados fossem cedidos à CA.
- Direito à Palavra-Violação do dever de não divulgação de publicações escritas.

Pela CA:

- Direito ao Caráter-Proibição de criação de perfis automatizados.
- Direito à Vida Privada-Proibição de tratamento de dados de filiação política
- Direito à Proteção de Dados-Não existiu uma preocupação de minimização de dados, de consentimento de tratamento de dados políticos
- Direito ao Esquecimento-Incapacidade de Apagar os dados completos de David Carroll(3.5.2)

### 3.5 Condenações

Apesar do mediatismo e dos muitos direitos potencialmente violados pelos envolvidos, as condenações foram escassas e as compensações bastante baixas para o que

seria esperado pela sociedade. As condenações às entidades são apresentadas a seguir:

#### 3.5.1 Ao Facebook

O Facebook, fruto da sua posição de fornecedor dos dados à CA foi condenado pela justiça inglesa ao pagamento de cinquenta mil libras por violar o RGPD[20].

#### 3.5.2 À CA: O Caso David Carroll

A única condenação que a Cambridge Analytica(CA) sofreu pelos atos praticados durante os anos de atividade ficou conhecido como o Caso David Carroll. David Carroll um professor de design norte-americano, decidiu pedir, ao abrigo do RGPD, o acesso aos dados que a CA tinha acerca de si.

Por ser incapaz de dotar David Carroll da totalidade da informação requisitada e em tempo oportuno, a empresa acabou condenada pela justiça do Reino Unido ao pagamento de dezoito mil libras[21].

Outros processos estavam em curso quando a empresa abriu falência. Este facto pôs termo a muitas ações legais em curso, sendo, portanto, o caso David Carroll a única condenação à CA transitada em julgada[22].

## 4 Proposta de Regulamento para a IA

Em 2021, a Comissão Europeia lançou uma proposta de regulamento europeu para tutelar juridicamente o desenvolvimento em IA[23]. Este documento visa "clarificar" e contribuir para tornar a UE uma "powerhouse" de desenvolvimento em IA.

### 4.1 Usos Proibidos de IA

A Comissão pretende clarificar neste documento as linhas vermelhas do que não é possível fazer com IA. A principal preocupação desta instituição é banir atividades de social scoring[24] e restringir o uso da biometria a casos muito particular, como procura de pessoas desaparecidas ou atos terroristas. O profiling

é, também novamente, reforçado como estritamente proibido como já é estipulado no RGPD[11].

## **4.2 Classificação de Atividades como IA de Alto Risco**

O documento pretende ainda classificar as atividades que usam IA, em críticas e não críticas. Entende-se por IA crítica, sistemas de IA integrados em plataformas críticas ao funcionamento das sociedades modernas (como centrais nucleares) e/ou atividades que sejam inerentemente mais intrusivas na vida dos cidadãos (uso de biometria). Quando classificados como IA crítica, os fabricantes devem obedecer a um caderno de encargos muito restrito, existindo um foco muito elevado na transparência e acesso de informação por parte das autoridades nacionais e europeias.

## **4.3 Promoção de Startups em IA: Uso de Sandboxings Legais**

A comissão pretende que Startups e PMEs de baixos recursos consigam utilizar sistemas de IA com facilidade[25]. Consagrando por isso a possibilidade destas usufruírem de "relaxamentos legais", vulgarmente definidas de sandboxes[26] por um período muito limitado e restrito ao teste e validação de modelos de negócio, sob requisição e autorização previamente do regulador.

## **4.4 Estado Atual da Proposta**

Após a consulta de diversas instituições europeias como o BCE ou o Conselho da Europa, em breve esta proposta irá ser votada no parlamento europeu, podendo este regulamento estar em vigor até 2024.

# **5 Conclusão**

Neste documento apresentou-se numa linguagem acessível a profissionais na área do IT uma perspetiva abrangente de quais as questões legais que se podem colocar durante o trabalho em NLP.

Dado se tratar de um documento escrito por um aluno de Engenharia Informática, ele não deve ser encarado como uma dispensa para a consulta de um jurista especializado. Deve ser antes visto como uma chamada de atenção para as questões legais que se podem colocar.

À medida que mais jurisprudência for produzida, os cidadãos terão uma ideia mais clara do que o legislador pretende nas muitas leis recentemente introduzidas nos campos da proteção de dados e do Machine Learning.

Contudo, ficou claro que é possível desenvolver ciência com liberdade de criação, desde que padrões éticos estejam sempre presentes no quotidiano dos investigadores.



## References

- [1] L. P. Reis, *Iart course presentation*, 2021.
- [2] *Artigo 35 da crp*, 2005. [Online]. Available: [http://bdjur.almedina.net/citem.php?field=item\\_id%5C&value=842628](http://bdjur.almedina.net/citem.php?field=item_id%5C&value=842628).
- [3] *Artigo 26 da crp*, 2005. [Online]. Available: [http://bdjur.almedina.net/citem.php?field=item\\_id%5C&value=842619](http://bdjur.almedina.net/citem.php?field=item_id%5C&value=842619).
- [4] T. Nobre, *Direito de autor e direitos conexos*, 2012. [Online]. Available: <https://www.ulisboa.pt/en/info/direito-de-autor-e-direitos-conexos>.
- [5] *Creative commons*, Jun. 2022. [Online]. Available: [https://pt.wikipedia.org/wiki/Creative\\_Commons](https://pt.wikipedia.org/wiki/Creative_Commons).
- [6] A. O. Thunström, *We asked gpt-3 to write an academic paper about itself-then we tried to get it published*, Jun. 2022. [Online]. Available: <https://www.scientificamerican.com/article/we-asked-gpt-3-to-write-an-academic-paper-about-itself-mdash-then-we-tried-to-get-it-published/>.
- [7] B. M. Mafalda, “Inteligência artificial, e-persons e direito: Desafios e perspetivas1,” *Revista Jurídica Luso-Brasileira N<sup>o</sup>6*, pp. 1477–1488, 2017.
- [8] *Artigo 79 do cc*. [Online]. Available: [http://bdjur.almedina.net/citem.php?field=item\\_id%5C&value=970464](http://bdjur.almedina.net/citem.php?field=item_id%5C&value=970464).
- [9] *Regime jurídico das instituições científicas*, 2019. [Online]. Available: <https://dre.pt/dre/detalhe/decreto-lei/63-2019-122317422>.
- [10] *Artigo 22 do rgpd*. [Online]. Available: <https://dre.pt/dre/lexionario/termo/deciso-es-individuais-automatizadas-rgpd>.
- [11] *Lex - 32016r0679 - en - eur-lex*. [Online]. Available: <https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=celex%3A32016R0679>.
- [12] M. R. Guimarães, *Protecção de dados*, Nov. 2022.
- [13] *Lex - 62012cj0131 - en - eur-lex*, 2014. [Online]. Available: <https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX%5C%3A62012CJ0131>.
- [14] dgsi, *Igfej - bases juridico-documentais*. [Online]. Available: <http://www.dgsi.pt/>.
- [15] *Lex acesso ao direito da união europeia*, 0AD. [Online]. Available: <https://eur-lex.europa.eu/homepage.html?locale=pt>.
- [16] *Cambridge analytica*, Aug. 2022. [Online]. Available: [https://pt.wikipedia.org/wiki/Cambridge\\_Analytica](https://pt.wikipedia.org/wiki/Cambridge_Analytica).
- [17] Wiki, *Ted cruz*, Jun. 2022. [Online]. Available: [https://pt.wikipedia.org/wiki/Ted\\_Cruz](https://pt.wikipedia.org/wiki/Ted_Cruz).
- [18] Wiki, *Donald trump*, Nov. 2022. [Online]. Available: [https://pt.wikipedia.org/wiki/Donald\\_Trump](https://pt.wikipedia.org/wiki/Donald_Trump).
- [19] *Uk independence party*, Oct. 2022. [Online]. Available: [https://en.wikipedia.org/wiki/UK\\_Independence\\_Party](https://en.wikipedia.org/wiki/UK_Independence_Party).
- [20] S. Reuters, *Facebook agrees to pay uk fine over cambridge analytica scandal*, Oct. 2019. [Online]. Available: <https://www.reuters.com/article/us-facebook-privacy-britain-idUSKBN1X9130>.

- [21] J. Wakefield, *Cambridge analytica taken to court over data storage*, Mar. 2018. [Online]. Available: <https://www.bbc.com/news/technology-43501184>.
- [22] B. News, *Cambridge analytica starts bankruptcy proceedings in us*, May 2018. [Online]. Available: <https://www.bbc.com/news/technology-44167000>.
- [23] *Lex - 52021pc0206 - en - eur-lex*. [Online]. Available: <https://eur-lex.europa.eu/legal-content/EN/HIS/?uri=CELEX%5C%3A52021PC0206>.
- [24] *Social credit system*, Dec. 2022. [Online]. Available: [https://en.wikipedia.org/wiki/Social\\_Credit\\_System](https://en.wikipedia.org/wiki/Social_Credit_System).
- [25] M. Cocco and I. Kindylidi, *Proposta de regulamento sobre a abordagem europeia para a inteligência artificial<sup>1</sup> - insights & media - vda - vieira de almeida, sociedade de advogados*. [Online]. Available: <https://www.vda.pt/pt/publicacoes/insights/proposta-de-regulamento-sobre-a-abordagem-europeia-para-a-inteligencia-artificial/23435/>.
- [26] B. -. F. M. for Economics Affairs and C. Action, *Regulatory sandboxes – testing environments for innovation and regulation*. [Online]. Available: <https://www.bmwk.de/Redaktion/EN/Dossier/regulatory-sandboxes.html>.

# Índice

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	O Processamento de Linguagem Natural . . . . .	1
1.2	Os Sucessos do NLP . . . . .	1
1.3	A Aprendizagem Computacional . . . . .	1
1.4	As Redes Neurais Artificiais Profundas . . . . .	1
1.5	O Processo de Treino de RNs . . . . .	2
1.6	Datasets: Uma Necessidade . . . . .	2
1.7	Organização do Documento . . . . .	2
<b>2</b>	<b>Direitos Conflitantes</b>	<b>2</b>
2.1	Direitos de Utilização Informática . . . . .	2
2.1.1	Reserva ao Tratamento de Dados Sensíveis . . . . .	2
2.2	Direito De Autor . . . . .	3
2.2.1	Direito Inato À Criação . . . . .	3
2.2.2	Direitos Consagrados . . . . .	3
2.2.3	Conteúdos Cobertos . . . . .	3
2.2.4	Conteúdos Isentos . . . . .	3
2.2.5	Licenças de Livre Acesso: A Creative Commons . . . . .	3
2.2.6	Dados Produzidos pela Máquina . . . . .	3
2.3	Direito à Palavra . . . . .	4
2.3.1	Semelhanças com o Direito à Imagem . . . . .	4
2.3.2	Diferença com Direitos de Autor . . . . .	4
2.3.3	Justificações da Violação . . . . .	4
2.3.4	”A Finalidade Científica” . . . . .	4
2.3.5	Um Conflito Grave com NLP . . . . .	4
2.4	Direito ao Carácter . . . . .	5
2.5	Direito à Proteção de Dados . . . . .	5
2.5.1	Conceito de Dados Pessoais . . . . .	5
2.5.2	Conceito de Tratamento de Dados Pessoais . . . . .	5
2.5.3	Direitos Consagrados . . . . .	5
2.5.4	Minimização dos Dados . . . . .	5
2.5.5	Localização Geográfica dos Dados . . . . .	5
2.5.6	Cobertura de Dados Automatizados-O Caso Google Spain . . . . .	6
2.5.7	Tratamento de Dados Para Investigação Científica . . . . .	6
2.6	Comentário à Escassez de Jurisprudência . . . . .	6
<b>3</b>	<b>Caso Prático: O Estranho Caso da Cambridge Analytica</b>	<b>6</b>
3.1	A Crítica Apontada . . . . .	6
3.2	Papel dos Envolvidos . . . . .	6
3.3	O Papel do NLP na CA . . . . .	7
3.4	Direitos Violados . . . . .	7

3.5	Condenações . . . . .	7
3.5.1	Ao Facebook . . . . .	7
3.5.2	À CA: O Caso David Carroll . . . . .	7
<b>4</b>	<b>Proposta de Regulamento para a IA</b>	<b>7</b>
4.1	Usos Proibidos de IA . . . . .	7
4.2	Classificação de Atividades como IA de Alto Risco . . . . .	8
4.3	Promoção de Startups em IA: Uso de Sandboxings Legais . . . . .	8
4.4	Estado Atual da Proposta . . . . .	8
<b>5</b>	<b>Conclusão</b>	<b>8</b>