DOI: 10.1590/s0103-4014.202438111.021

Democracia aumentada: Inteligência Artificial como ferramenta de combate à desinformação

Alexandre Alcoforado, [I]
Thomas Palmeira Ferraz, [II] *Enzo Bustos,* [III]
André Seidel Oliveira, [IV] *Rodrigo Gerber,* [V]
Gian Lucca du Mont Santoro, [VI]
Israel Campos Fama, [VII] *Bruno Miguel Veloso,* [VIII]
Fábio Levy Siqueira, [IX] *Anna Helena Reali costa* [X]

Introdução

Cidadãos bem-informados e espaços de debate e crítica são princípios fundamentais da democracia. Sem eles, a qualidade do debate pelo qual as sociedades devem passar para melhorar a si mesmas pode degradar (Prothro; Grigg, 1960).

Na última década, houve uma mudança significativa no conceito de espaço de informação e discussão: as novas tecnologias permitiram que as pessoas se organizassem para expressar as suas opiniões e mudar os regimes políticos dos seus países. Um exemplo instigante do fenômeno foi a "Primavera Árabe", em 2011, quando mobilizações promovidas nas redes sociais, como YouTube, Facebook e Twitter, causaram uma grande mobilização dos cidadãos (Safranek, 2012). No entanto, essas mesmas tecnologias têm gerado tensão social e polarização, pois facilitam a disseminação de notícias falsas e discursos extremistas e podem até ter influência nos resultados eleitorais (Tucker et al., 2018). Por exemplo, em 8 de janeiro de 2023, milhares de pessoas invadiram as sedes dos três Poderes da República em Brasília. Nos meses que antecederam a invasão, mídias sociais foram bombardeadas com notícias falsas que, além de colocar em dúvida a lisura do processo eleitoral brasileiro, aventavam uma suposta legitimidade constitucional para que as Forças Armadas assumissem o poder (Mota, 2023). Portanto, as tecnologias nos apresentam um desafio e uma oportunidade: combater a disseminação de desinformação; e utilizar esse novo meio como um serviço à democracia para aproximar a política dos cidadãos e promover o debate sobre as principais questões e importantes decisões. Este desafio ficou ainda maior com

o lançamento do ChatGPT[1] em 2022, um modelo de linguagem desenvolvido pela OpenAI, capaz de gerar textos a partir de um *prompt* fornecido pelo usuário. Ou seja, além de potencializar o alcance de notícias falsas, a tecnologia agora é capaz de criar conteúdo (textos bem estruturados que, por vezes, são difíceis de distinguir se foram gerados por humanos ou máquinas).

Vale ressaltar que a adoção dessas tecnologias mudou aspectos da vida em sociedade. Entre as muitas razões para isso, a popularização de dispositivos conectados remotamente permite que as pessoas consumam informações e publiquem suas próprias opiniões com extrema rapidez. No entanto, embora a vida tenha se simplificado em muitos aspectos, a maioria dos aspectos da política continua sendo conduzida da mesma forma, sem explorar o poder de síntese presente na maioria dos conteúdos digitais. O que ocorre, portanto, é um descompasso entre os cidadãos – e sua vida digital – e o comportamento político de seus países e do mundo, o que pode ser um dos fatores que contribuem fortemente para a atual redução da crença e confiança nas instituições democráticas (Simon et al., 2017). Nesse contexto, ganha força o conceito de *e-democracia* ou democracia digital: pode ser entendida como o uso da tecnologia no processo de formulação de políticas e nas relações cidadão-estado por meio da criação de ferramentas que estimulem a participação direta do cidadão nas decisões e discussões que a sociedade deve passar (Council of Europe, 2009). Utilizando essas ferramentas, os cidadãos podem ser muito mais ativos na vida pública e no processo de tomada de decisão (Breindl; Francq, 2008). Vedel (2003) define três eixos para a democracia digital: informação, discussão e decisão. Podemos interpretar esses eixos como a forma como o cidadão pode agir politicamente na Internet: informar-se, ser ouvido no debate e participar efetivamente das decisões.

Por outro lado, a desinformação pode ser um problema central para a democracia digital, pois cidadãos desinformados produzem discussões e decisões ruins. Esse fenômeno ocorre fundamentalmente porque não ter informação ou ter informação falsa causa desinformação. No entanto, o excesso de informação (Bontcheva; Gorrell; Wessels, 2013) é também uma causa fundamental da desinformação, muito agravada por este contexto digital em que surgiram as redes sociais. A informação disponível nessas redes é ruidosa no sentido de que há muito mais dados do que qualquer pessoa pode consumir. Pela sua própria natureza, as redes sociais têm provocado um consumo mais passivo de informação na Internet: em vez de a procurar ativamente, as pessoas passam a maior parte do tempo a filtrar e a gerir a enorme quantidade de informação que recebem diariamente.

É importante ressaltar que as informações produzidas pelos órgãos estatais também são ruidosas em geral: muitas vezes consistem numa extensa documentação de difícil interpretação para os cidadãos, facilitando que grupos maliciosos utilizem partes descontextualizadas desses dados para enganar a população. Esse excesso de dados produzidos a partir de documentos de órgãos estatais pode ser caracterizado como *Big Data*, pois são ativos de informação com alto volume,

alta velocidade e alta variedade (ou os três "V") que requerem técnicas e tecnologias avançadas e inovadoras para captura, armazenamento e processamento das informações para uma melhor análise e tomada de decisão (Gandomi; Haider, 2015). Portanto, como o conjunto de todas as informações geradas pelos órgãos estatais pode ser caracterizado como *Big Data*, a criação de serviços e ferramentas para a divulgação, de forma clara, direta e objetiva, de materiais que estimulem a participação dos cidadãos nas discussões políticas em curso torna-se extremamente importante.

Em busca de melhores formas de lidar com *Big Data*, notamos que a última década proporcionou uma revolução na Inteligência Artificial, expandindo horizontes e trazendo descobertas de novas tarefas que as máquinas podem realizar. Em particular, as técnicas de processamento de linguagem natural já têm o poder de interpretar automaticamente grandes volumes de informações (Young et al., 2018). Essa capacidade permite, por exemplo, extrair informações relevantes e diretas do grande número de documentos produzidos por órgãos estatais, muitos dos quais compostos por textos longos e complexos. Criar uma ferramenta que utiliza essas técnicas para simplificar a informação e traduzi-la para o público teria muitas aplicações práticas, entre as quais citamos: auxiliar as agências de notícias na consulta de informações e na elaboração de trabalhos jornalísticos baseados em dados (*Data Journalism*), e fornecer um aumento da transparência das ações da classe política perante a sociedade. Portanto, há uma oportunidade de aplicar técnicas de processamento de linguagem natural para *Big Data* produzido por agências estatais.

Nesse contexto, este trabalho visa contribuir com uma ferramenta que permite a coleta, organização, seleção, processamento e simplificação automática das informações produzidas nas discussões políticas. A ferramenta reduz o ruído de dados para obter melhores informações processando os dados públicos disponíveis. Assim, abordamos um dos eixos essenciais da democracia digital, que é a informação. Em particular, mostramos um estudo de caso para validar nossa ferramenta de democracia aumentada neste trabalho. Recorremos à Assembleia da República e utilizamos as atas das reuniões plenárias descritas no Diário da Assembleia da República Portuguesa, procurando desenvolver uma ferramenta que ajude a clarificar o trabalho legislativo. Nossa principal contribuição é demonstrar o potencial e a viabilidade da ferramenta proposta que:

1 Processa a massiva e complexa quantidade de dados produzidos no corpo legislativo, segmentando as atas em peças de discussão;
2 Extrai conhecimento do discurso político, identificando atores, temas, assuntos de interesse, projetos e citações nas discussões;
3 Resume (ou seja, comprime) informações relevantes, transformando grandes *corpora* de texto em um texto representativo menor e fácil de ler;
4 Fornece um conjunto de *dashboards* (painéis de análise) e informações para cidadãos e organizações.

O restante deste artigo está organizado da seguinte forma: inicialmente, descrevemos trabalhos relacionados existentes, buscando inspiração para funcionalidades e detetando semelhanças e diferenças com a nossa proposta. Em seguida, descrevemos a nossa ferramenta proposta e sintetizamos as suas funcionalidades, que são descritas nas seções seguintes e ilustradas com exemplos representativos retirados do estudo de caso realizado com as atas da Assembleia da República. Por fim, encerramos com nossas discussões e propostas para trabalhos futuros.

Trabalhos relacionados

Alguns autores já tratam da análise textual de documentos públicos. A maioria das aplicações desenvolvidas nesse contexto está na classe de análise de sentimentos, um ramo do processamento de linguagem natural. A maioria dos trabalhos na literatura tratou do idioma inglês, mas pesquisas recentes em outros idiomas surgiram com mais frequência (Rao, 2019). O objetivo da análise de sentimentos é classificar a opinião expressa em textos como sendo positiva, neutra ou negativa sobre os itens comentados.

Um dos trabalhos pioneiros foi o desenvolvido por Mullen e Malouf (2006). Eles exploraram a análise de sentimentos do discurso político informal, relatando testes estatísticos realizados em dados de um grupo de discurso político americano em inglês. Embora os resultados tenham sido preliminares, fica claro que métodos simples de classificação de textos nem sempre fornecem resultados satisfatórios para dados tão complexos quanto textos de discurso político.

Porém, nos últimos anos, o processamento de linguagem natural passou por uma transformação significativa devido a poderosas técnicas de Inteligência Artificial, como redes neurais e *Transformers*, possibilitando resultados mais empolgantes. A maioria dessas técnicas, no entanto, segue um paradigma de aprendizagem supervisionada que requer uma grande quantidade de dados rotulados para treiná-las, o que nem sempre está disponível.

Recentemente, Watanabe e Zhou (2022) aplicaram técnicas semissupervisionadas de classificação de texto em documentos das Nações Unidas. Técnicas semissupervisionadas visam reduzir a dependência de dados rotulados no treinamento. Elas alcançaram resultados robustos nos idiomas japonês e inglês usando no treinamento do algoritmo apenas algumas poucas palavras que refletem sentimentos. Embora os resultados tenham sido promissores, ainda são necessárias palavras-chave para classificação, o que pode ser uma dificuldade em potencial. É do nosso interesse explorar métodos de classificação que necessitem apenas de uma lista de classes possíveis, não dependendo de outras informações auxiliares fornecidas pelo usuário.

Um exemplo inspirador para o nosso trabalho é o projeto "Decide Madrid" (Procter et al., 2021), que, além de resumir informações políticas, permite que os cidadãos opinem sobre os projetos em votação. Entre outros objetivos,

o projeto "Decide Madrid" visa identificar as posições políticas dos cidadãos, agrupar os cidadãos por interesses semelhantes e sugerir propostas que as pessoas possam querer apoiar. O projeto explora intensamente um ramo da Inteligência Artificial que é a aprendizagem de máquina, assim como nós fazemos neste trabalho. No entanto, no estado atual de nossa ferramenta, a plataforma oferecida pelo projeto "Decide Madrid" é muito mais interativa com quem a utiliza, enquanto a nossa ferramenta oferece mais visualização de dados.

No Brasil, Silva et al. (2021) aplicam a modelagem de tópicos a um conjunto de dados fornecido pela Câmara de Inovação e Tecnologia da Informação da Câmara dos Deputados, obtendo resultados robustos e mostrando a eficácia da aplicação de técnicas modernas de processamento de linguagem natural ao discurso político. Essa obra nos inspira porque instiga a curiosidade de ver o que acontece quando não uma, mas muitas técnicas são aplicadas ao discurso político.

Apesar de não ser o foco da ferramenta proposta, é interessante observar que a quantidade de dados gerados não representa um desafio apenas para o cidadão comum. Técnicos e agentes políticos das mais diversas instâncias do poder público são demandados a analisar quantidades crescentes de dados em suas rotinas de trabalho. No ramo jurídico, por exemplo (Carmo et al., 2023) apontam a existência de cerca de 77,3 milhões de processos em tramitação no sistema judicial brasileiro em 2022; com um volume tão expressivo, o uso de tecnologias capazes de dar maior celeridade ao fluxo processual tem relevante impacto positivo na sociedade. Neste sentido, os autores citam o Programa Justiça 4.0; uma iniciativa governamental que busca promover soluções digitais para modernizar o Poder Judiciário brasileiro.

No cenário português, existe um aplicativo móvel chamado meuparlamento.pt[2] que permite às pessoas verificarem a que partido político estão mais próximas quando são analisadas apenas as votações realizadas em projetos específicos. Esse aplicativo aborda o mesmo domínio da nossa ferramenta, a Assembleia da República; no entanto, não há exploração de técnicas de Inteligência Artificial para extrair conhecimentos mais elaborados das atas das sessões parlamentares – e esse é o nosso objetivo nesse trabalho.

Quinta-feira, 17 de Setembro de 1998 I Série — Número 1

DIÁRIO
da Assembleia da República

VII LEGISLATURA 4.ª SESSÃO LEGISLATIVA (1998-1999)

REUNIÃO PLENÁRIA DE 16 DE SETEMBRO DE 1998

Presidente: Ex.mo Sr. António de Almeida Santos

Secretários: Ex.mos Srs. Artur Rodrigues Pereira dos Penedos
Duarte Rogério Matos Ventura Pacheco
João Cerveira Corregedor da Fonseca
Rosa Maria da Silva Bastos da Horta Albernaz

SUMÁRIO

O Sr. Presidente, que saudou os Deputados na abertura do ano parlamentar, declarou aberta a sessão às 15 horas e 25 minutos.

Antes da ordem do dia. — *Deu-se conta da entrada na Mesa da proposta de lei n.º 207/VII, da proposta de resolução n.º 119/VII e do projecto de lei n.º 556/VII, bem como de requerimentos e da resposta a alguns outros.*

Em declaração política, o Sr. Deputado Francisco de Assis (PS) elogiou a acção governativa em vários domínios e criticou o PSD pela sua postura em relação não só à regionalização, mas também ao inquérito parlamentar realizado na sequência das acusações feitas ao Governo pelo líder do Partido Social Democrata, no último Congresso do PSD realizado no Algarve, e por este partido ter apresentado um projecto de lei que tem a ver com a questão da Sport TV.

Também em declaração política, o Sr. Deputado Carlos Encarnação (PSD) acusou o Sr. Ministro da Agricultura por ter desprezado o protesto dos agricultores, assim como o Sr. Primeiro-Ministro pelo modelo de regiões apresentado pelo PS.

Igualmente em declaração política, o Sr. Deputado Octávio Teixeira (PCP) chamou a atenção para o adiamento por parte do Governo do cumprimento de promessas eleitorais.

Por fim, também em declaração política, o Sr. Deputado Nuno Abecasis (CDS-PP) fez um apelo aos Deputados portugueses de solidariedade para com os representantes da UNITA no Parlamento angolano.

A requerimento do PS, a discussão e votação do voto n.º 130/VII — De protesto pela falta de resposta do Governo aos problemas da lavoura nacional (CDS-PP) foram adiadas para a sessão plenária do dia seguinte. Usaram da palavra os Srs. Deputados Acácio Barreiros (PS) e Luís Queiró (CDS-PP).

A Câmara aprovou um relatório e parecer da Comissão de Assuntos Constitucionais, Direitos, Liberdades e Garantias sobre a retoma de um Deputado do PSD e a substituição de três Deputados, um do PS, outro do PCP e outro do PSD.

Ordem do dia. — *Procedeu-se à apreciação conjunta das propostas de resolução n.os 106/VII — Aprova, para ratificação, o Protocolo de Adesão da República da Polónia ao Tratado do Atlântico Norte, assinado em Bruxelas, em 16 de Dezembro de 1997, 107/VII — Aprova, para ratificação, o Protocolo de Adesão da República Checa ao Tratado do Atlântico Norte, assinado em Bruxelas, em 16 de Dezembro de 1997, e 108/VII — Aprova, para ratificação, o Protocolo de Adesão da República da Hungria ao Tratado do Atlântico Norte, assinado em Bruxelas, em 16 de Dezembro de 1997, tendo as mesmas sido aprovadas em votação global. Intervieram, além do Sr. Ministro dos Negócios Estrangeiros (Jaime Gama), os Srs. Deputados Pedro Holstein Campilho (PSD), Laurentino Dias (PS), Nuno Abecasis (CDS-PP) e João Amaral (PCP).*

O projecto de lei n.º 530/VII — Privatização do notariado (PSD) foi discutido na generalidade, tendo usado da palavra, a diverso título, os Srs. Deputados Maria Eduarda Azevedo (PSD), Odete Santos (PCP), Nuno Baltazar Mendes (PS), Guilherme Silva (PSD) e Silvio Rui Cervan (CDS-PP).

O Sr. Presidente encerrou a sessão eram 18 horas e 40 minutos.

Fonte: Sítio da Assembleia da República, DAR I Série n.001 (acesso em: 13 jan. 2023).

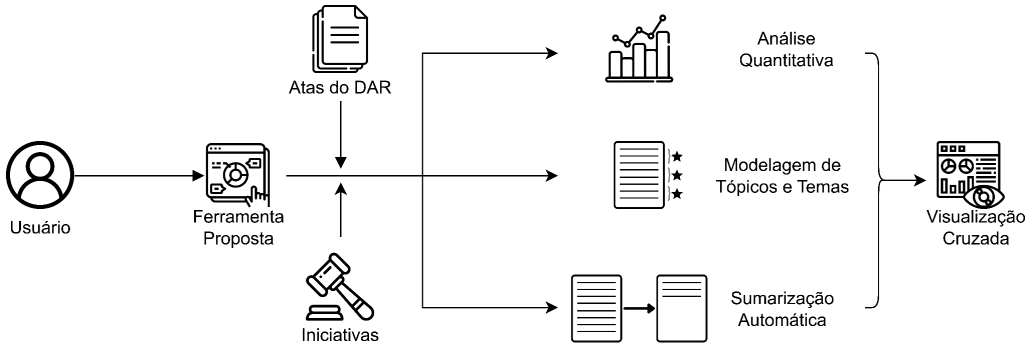
Figura 1 – "Diário da Assembleia da República" (DAR).

Ferramenta proposta

Portugal é uma democracia desde 1974 e, em 1975, começaram a ser transcritas as atas da Assembleia da República. A partir de 1976, na 3ª República, surgiram as transcrições do Diário da Assembleia da República (DAR), ilustradas na Figura 1, disponíveis gratuitamente ao público (Assembleia da República, 2023). As transcrições das atas do DAR são utilizadas pela ferramenta proposta para extrair informações de interesse público.

A ferramenta proposta neste trabalho pressupõe determinar quais aspectos da informação são relevantes para o público e apresentar os resultados de forma simples e visual em uma interface de fácil manuseio. Ao centrar-nos na Assembleia da República, procuramos identificar quais informações poderiam interessar aos cidadãos e quais as informações já produzidas pelas atas do DAR poderiam ser resumidas e expostas ao cidadão de formas diferentes e compreensíveis.

A Figura 2 ilustra o esquema geral da ferramenta proposta e suas respectivas funcionalidades. Numa primeira fase, tanto as atas do DAR como as iniciativas discutidas no parlamento são acessadas, e as transcrições do parlamento são retiradas do site, onde todas as atas estão disponíveis publicamente e gratuitamente. Os textos das atas são previamente processados e armazenados em uma base de dados, que então ficará disponível para as funcionalidades acessarem, conforme descrito na próxima seção. A partir das atas e iniciativas do DAR obtidas na etapa de coleta de dados e armazenadas no banco de dados, a ferramenta pode computar automaticamente as citações faladas pelos parlamentares nas sessões de interesse, classificar as discussões das sessões em tópicos e temas, e resumir partes das atas. Cada uma dessas funcionalidades e respectivas formas de apresentação na interface da ferramenta são descritas nas seções a seguir.



Fonte: Autoria própria.

Figura 2 – Esquema geral da ferramenta proposta e suas respectivas funcionalidades: (i) cálculo de citações diretas ou indiretas; (ii) classificação dos tópicos e temas tratados nas atas; e (iii) sumarização automática das partes de interesse da ata. Essas funcionalidades acessam um banco de dados gerado a partir das atas do DAR e das iniciativas discutidas nas sessões parlamentares. Os resultados são mostrados em uma interface com o usuário.

Coleta de dados

Inicialmente, foram utilizadas técnicas de *web crawling* (Pant; Srinivasan; Menczer, 2004) e *web scraping* (Mitchell, 2018) para recolher dados publicados periodicamente no sítio da Assembleia da República, nomeadamente as atas das sessões parlamentares (Assembleia da República, 2021). Esses dados estão disponíveis na forma de textos, cada um contendo as transcrições de uma sessão. Em seguida, estruturou-se o conteúdo desses documentos em um banco de dados para facilitar o processamento, construindo um grande *corpus* de texto. Esses dados foram primeiramente segmentados nos itens da pauta de cada reunião, que envolvem declarações políticas, discussões e votações de iniciativas parlamentares (projetos de lei, projetos de resolução, consultas parlamentares, entre outros), além de outros assuntos. Essas tarefas são referidas como a etapa de coleta de dados.

Com essa base de dados organizada, foram aplicadas técnicas de aprendizagem de máquina associadas a métodos de processamento de linguagem natural para fornecer as funcionalidades ao usuário. Todos os resultados aqui relatados foram obtidos através da realização de experiências com atas de 16.9.2020 a 25.2.2021.

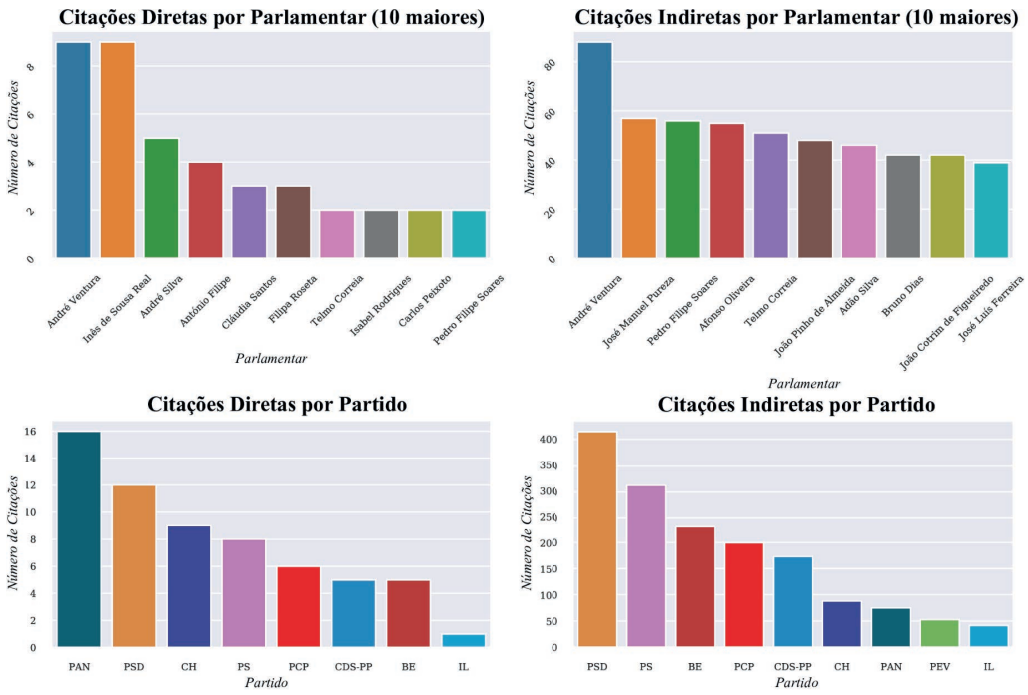
Computação das citações

Definimos *assunto de interesse* como um conjunto de palavras-chave definidas pelo usuário relacionadas a um assunto, como "corrupção" ou "educação". Definimos *citações diretas* como discursos de parlamentares que mencionam explicitamente as palavras-chave. As demais falas contidas na mesma discussão em que houve pelo menos uma citação direta são *citações indiretas*. Uma funcionalidade apresentada pela ferramenta de democracia aumentada proposta neste trabalho utiliza dados segmentados das declarações políticas para, de acordo com o assunto de interesse, realizar cálculos de citações diretas e indiretas para cada partido e parlamentar dentro do intervalo de tempo definido pelo usuário.

Para possibilitar a busca de citações diretas e indiretas, foi aplicado nas atas o algoritmo DEBACER (Ferraz et al., 2021) para particionar os conjuntos de todas as declarações políticas de cada sessão parlamentar em blocos de fala que contenham discussões completas (com um início e fim), o que é essencial para a compreensão do que está sendo dito pelos participantes. Armazenamos essas partições numa base de dados e desenvolvemos um procedimento de análise. Palavras-chave definidas no assunto de interesse foram então pesquisadas nessas partições, identificando o parlamentar que as citou explicitamente. A partir desses dados, várias informações podem ser geradas. Em particular, a ferramenta mostra com que frequência cada parlamentar (ou partido) cita a palavra-chave no período de interesse definido pelo usuário. Citações indiretas, tanto de um membro do parlamento quanto de partidos, também são computadas.

Os resultados de um estudo de caso para o assunto de interesse "corrupção" é apresentado para mostrar a eficácia da ferramenta proposta, analisando a participação de partidos e parlamentares neste assunto. A importância do exemplo da corrupção no caso português pode ser fundamentada por vários fatos (Prémio Tágides, 2021), tais como (i) o custo estimado para os portugueses de casos de corrupção conhecidos equivale a 30 % da dívida pública nacional; (ii) o fato de apenas 1 das 15 leis anticorrupção recomendadas em 2016 ter sido integralmente implementada em Portugal; e (iii) também o fato de o Parlamento Europeu ter estimado que a corrupção em Portugal custa o equivalente a 8 % a 10 % do PIB.

A Figura 3 apresenta um exemplo de informação que é disponibilizada ao usuário. A figura mostra os histogramas que indicam o número de citações diretas e indiretas da palavra "corrupção" definida pelo usuário e seus derivados (por exemplo, corruptos, corrupções, etc.), calculados no período entre 16.9.2020 e 25.2.2021, seja por um deputado ou pelo partido.



Fonte: Autoria própria.

Figura 3 – Histogramas de citações diretas e indiretas derivadas da palavra "corrupção" no período de 16.9.2020 a 25.2.2021, por parlamentar (apenas os dez mais citados) e por partido.

É possível combinar esse recurso com outras funcionalidades. Por exemplo, poderíamos determinar o tema dominante sobre o qual a palavra-chave foi citada (por exemplo, "corrupção na saúde") se a funcionalidade de citação fosse combinada com aquela que descobre tópicos e temas, descrita na próxima seção.

Tópicos e temas

Uma segunda funcionalidade desenvolvida na ferramenta realiza duas etapas em atas de sessões parlamentares ou partes delas, de acordo com o período de interesse especificado. O primeiro passo é a modelagem de tópicos e o outro, a classificação dos discursos em temas.

Um *modelo de tópico* é um modelo estatístico de mineração de texto usado para descobrir estruturas semânticas ocultas em uma coleção de documentos. Definimos *tópico* como um conjunto de palavras que pertencem ao mesmo campo semântico, como "escola, professores, turma" ou "5G, ciência, rede". É utilizada a *modelagem de tópicos* (Blei; Ng; Jordan, 2003; Grootendorst, 2022) para determinar automaticamente os tópicos que ocorrem nos documentos. Ao contrário dos assuntos de interesse definidos pelo usuário, dados por palavras-chave usadas no cálculo das citações, os tópicos são extraídos automaticamente dos documentos por um algoritmo de Inteligência Artificial treinado em um idioma específico (aqui, o português).

O algoritmo de modelagem de tópicos encontrou vários tópicos e associou cada fala a eles. O algoritmo devolve a probabilidade que indica a relevância de cada tópico para o discurso. Por causa disso, os tópicos podem ser apresentados de maneiras diferentes. Por exemplo, mostrar o tópico com maior probabilidade de associação ao texto, ou os três principais tópicos, ou todos os tópicos e suas respectivas probabilidades e assim por diante.

No entanto, os tópicos encontrados pelos algoritmos de modelagem de tópicos nem sempre são interpretáveis por humanos. Eles formam agrupamentos de palavras, mas não possuem um nome próprio para cada grupo, dificultando a associação daquele agrupamento a um único significado. Assim, a funcionalidade da ferramenta permite que o usuário escolha se está satisfeito com esses tópicos encontrados automaticamente ou se ele prefere utilizar a segunda etapa da funcionalidade, que é classificar as falas em *temas*.

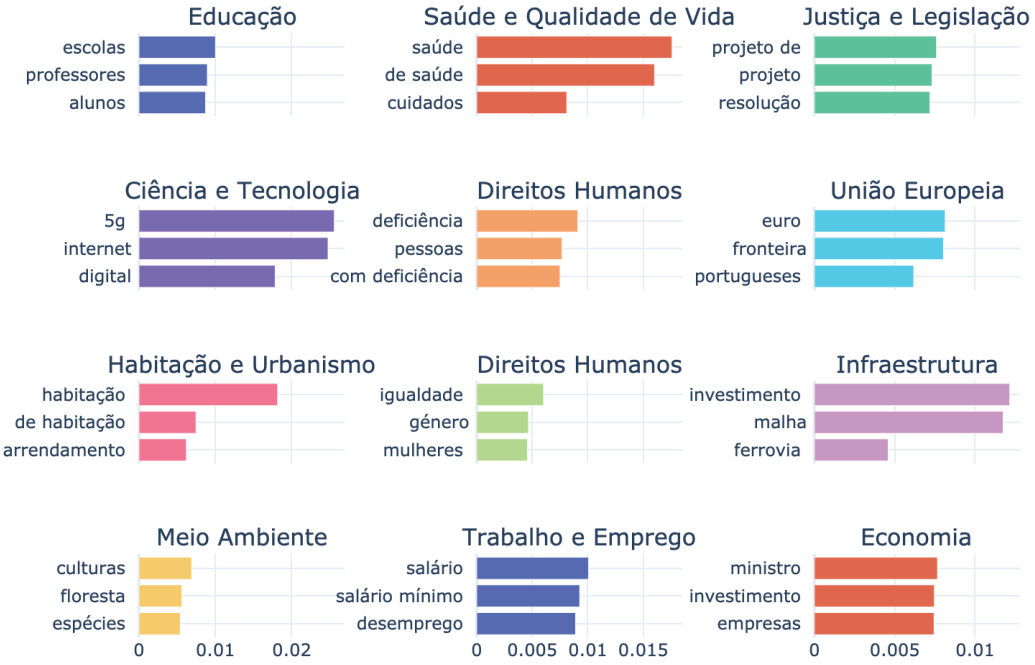
Os temas são definidos *a priori* na ferramenta e descrevem categorias de interesse da atividade parlamentar, como Economia, Meio Ambiente etc. Para classificar as falas em temas pré-estabelecidos, essa funcionalidade usa um algoritmo conhecido como Zero-Shot (Romera-Paredes; Torr, 2015; Yin; Hay; Roth, 2019). Tecnicamente, foi aplicado o algoritmo ZeroBERTo (Alcoforado et al., 2022), que considera: (i) os tópicos encontrados pelo algoritmo de modelagem de tópicos na etapa anterior; (ii) a distribuição de probabilidades de cada fala pertencer a cada tópico; (iii) os temas definidos previamente. Os temas são então associados probabilisticamente às falas por um modelo de linguagem treinado, potencializando seu conhecimento geral da língua. Semelhante à associação anterior da fala ao tópico, esta etapa associa cada fala a todos os temas previamente definidos. Assim, os temas também podem ser apresentados de diferentes maneiras. Por exemplo, o tema com maior probabilidade de associação ao texto, os cinco principais temas, todos os temas e suas respectivas probabilidades, um gráfico com a evolução temporal de um determinado tema, entre outros.

Um aspecto importante dessa funcionalidade é que as mesmas palavras-chave podem ser escolhidas tanto como assunto de interesse quanto como tema. No entanto, existem diferenças fundamentais entre eles:

- Primeiro, a entrada para o assunto de interesse na primeira funcionalidade é um único termo (por exemplo, "corrupção") e a partir dele é definido um conjunto de palavras-chave que derivam do termo dado (por exemplo, corrompe, corrompido, corrupção, etc.), enquanto a entrada para temas é necessariamente um conjunto de palavras relacionadas aos assuntos discutidos no parlamento (por exemplo, ciência e tecnologia, meio ambiente, direitos humanos, saúde e infraestrutura);
- Em segundo lugar, o assunto de interesse na primeira funcionalidade é pesquisado dentro do intervalo de tempo definido pelo usuário, procurando correspondências exatas ou aproximadas para essas palavras-chave, enquanto os temas são automaticamente associados aos discursos por um modelo de linguagem;

- Em terceiro lugar, um assunto de interesse pode ou não ser encontrado em uma fala em um determinado intervalo de tempo, enquanto cada tema, por sua vez, está associado a todas as falas, e a força dessa associação é representada por um valor de probabilidade, que pode ser semelhante para mais de um tema, isto é, uma determinada fala pode estar fortemente relacionada tanto ao tema Saúde quanto ao tema Direitos Humanos.

As informações sobre o modelo de tópico de um documento e a classificação do tema são exibidas visualmente na nossa ferramenta, conforme ilustrado na Figura 4. Na figura, vemos alguns exemplos apresentados na interface gráfica: os 12 tópicos mais frequentes encontrados, descritos por suas três palavras mais representativas (ex. o tópico rosa é definido pelas palavras "Investimento", "Malha" e "Ferrovia". Acima dos 12 tópicos apresentados, vemos os temas sobre os quais cada tópico foi melhor avaliado (por exemplo, o tópico rosa foi classificado primeiro como "Infraestrutura". Esses dados são referentes à ata do período de 16.9.2020 a 25.2.2021.



Fonte: Autoria própria.

Figura 4 – Tópicos encontrados (apenas os 12 tópicos mais frequentes), juntamente com suas palavras mais representativas (apenas as três mais representativas). Acima das barras está o tema associado a cada tópico específico (apenas o tema mais representativo), de 16.9.2020 a 25.2.2021.

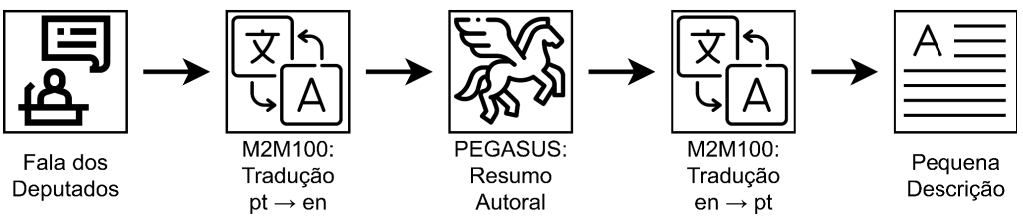
A aplicação dessa funcionalidade nas referidas atas resultou em cerca de 60 tópicos na etapa de modelagem de tópicos. Esta experiência foi realizada com o

seguinte conjunto de temas: Corrupção, Cultura, Economia, Educação, Energia, Meio Ambiente, União Europeia, Saúde e Qualidade de Vida, Habitação e Urbanismo, Direitos Humanos, Indústria e Agricultura e Comércio, Infraestrutura, Justiça, Legislação, Defesa Nacional e Segurança Pública, Ciência e Tecnologia, Turismo, Trabalho e Emprego.

Com essa funcionalidade, também é possível calcular, por exemplo, quantas vezes um determinado parlamentar ou partido falou sobre determinado tema. Outra opção é, por exemplo, resumir falas que tenham grande probabilidade de serem associadas a um tema específico de acordo com o desejo do usuário. A funcionalidade de resumo é descrita na próxima seção.

Sumarização

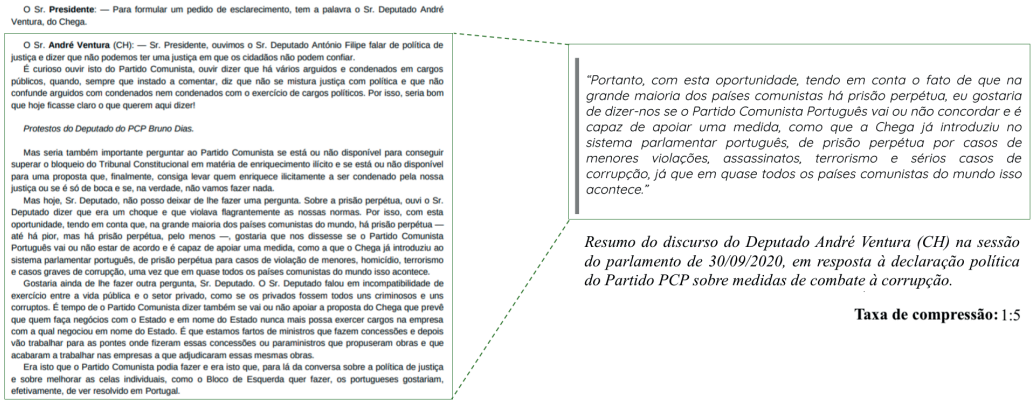
A funcionalidade de sumarização automática aplica redes neurais treinadas para sumarização com o objetivo de gerar relatórios curtos dos discursos relacionados a relatórios de iniciativas, em particular projetos de lei. Essas falas são mais longas que as demais e ocorrem logo após o início de uma discussão, com o objetivo de contextualizar a discussão para os demais parlamentares. Mais especificamente, aplicamos o modelo Pegasus (Zhang et al., 2020), uma rede neural *Transformer* (Vaswani et al., 2017) capaz de gerar resumos autorais em inglês, ou seja, resumos com frases originais que não são simples fragmentos dos textos fonte. Para fazer isso, o algoritmo utilizado pela ferramenta primeiro traduz o texto de origem do discurso para o inglês, produz o resumo com o Pegasus e, finalmente, traduz o resumo de volta para o português. É utilizado o tradutor M2M100 (Fan et al., 2021), pois é um modelo multilingue robusto baseado no *Transformer*. A Figura 5 ilustra o procedimento completo de resumo.



Fonte: Autoria própria.

Figura 5 – O resumo de um relatório de iniciativa é gerado a partir de textos traduzidos inicialmente do português para o inglês pelo tradutor M2M100, seguido pelo sumarizador inglês do PEGASUS e finalmente traduzidos novamente para o português, gerando resumos curtos.

Esses pequenos resumos das falas dos relatores de cada iniciativa são apresentados aos usuários como um relato das principais discussões ocorridas nas atas. Os usuários também podem usar a funcionalidade para gerar resumos adicionais da discussão completa sobre aquele projeto, ou podem ser redirecionados para as atas originais que transcrevem toda a discussão de interesse. Uma ilustração do texto original e seu resumo gerado está na Figura 6.



Fonte: Autoria própria.

Figura 6 – Exemplo de resumo: à esquerda está a parte de uma ata que foi resumida na figura à direita.

Alternativas para visualização

As funcionalidades apresentadas anteriormente produzem saídas como resumos de documentos, número de citações diretas ou indiretas de um determinado assunto de interesse e seus respectivos oradores, e os tópicos e temas abordados nas diferentes partes de interesse das atas das sessões parlamentares.

A interação das informações extraídas pelas diferentes funcionalidades da ferramenta desenvolvida permite muitas outras opções de visualização dos dados. Como todos os dados também estão disponíveis para filtros de intervalo de tempo e para partes de atas de interesse do usuário final, a ferramenta pode ser utilizada de forma personalizada.

Observa-se que informações externas também podem ser utilizadas em combinação com as informações produzidas pela ferramenta. Uma forma de visualização poderia ser, por exemplo, um gráfico comparativo entre o crescimento do PIB com a evolução temporal da discussão sobre o tema "Economia" que tenha ocorrido no parlamento, ou ainda entre a taxa de desemprego e o valor do salário mínimo com a evolução temporal do tema "Trabalho e Emprego".

Essas diferentes formas de visualizar dados ainda estão em desenvolvimento na nossa ferramenta. Após o desenvolvimento, ainda serão necessários testes de usabilidade com os usuários finais.

Discussões e trabalhos futuros

Neste artigo foi proposta uma ferramenta com uma interface gráfica interativa que simplifica a informação produzida pela Assembleia da República no intervalo de tempo definido pelo usuário e apresenta os seguintes resultados:

1 Estatísticas sobre a participação de um parlamentar ou partido em determinado assunto de interesse, como citações diretas e indiretas do assunto;

2 Os temas mais ou menos relevantes discutidos pelos deputados e partidos no período definido;

3 Os temas mais ou menos frequentes que correspondem aos tópicos extraídos das discussões;

4 Resumos de fala;

5 Estatísticas sobre discursos de deputado ou partido significativamente associados aos temas definidos.

Assim, a ferramenta proposta colabora diretamente com o eixo da informação, que é o primeiro eixo da democracia digital. Foram coletados e estruturados dados públicos, os quais foram processados, extraindo conhecimento que de outra forma seria impossível de obter devido à enorme quantidade de texto. Ao sintetizar, resumir, modelar e classificar esses dados disponíveis com técnicas de aprendizagem de máquina, mostrou-se que é possível fornecer formas de visualizar informações já públicas – mas não processadas – mais fáceis de entender pelos cidadãos e pela sociedade em geral. Além disso, a interação entre as diversas funcionalidades da ferramenta pode oferecer diversas opções de visualização. Diferentes possibilidades de visualização ainda estão sendo discutidas pelos autores com diferentes públicos de interesse.

Acreditamos que a informação produzida pela ferramenta permite aos cidadãos identificar o comportamento político dos deputados e partidos, avaliando a adesão às suas ideologias políticas, a coerência entre discurso e prática dos partidos e deputados. Como resultado, as pessoas podem incentivar uma maior atenção dos atores políticos às questões que a sociedade considera mais importantes.

A ferramenta proposta ainda está em desenvolvimento. Para avaliar sua viabilidade e o valor de seus resultados, foi realizada uma série de experiências em uma quantidade modesta de atas. Os algoritmos envolvidos nessas experiências precisam ser integrados à ferramenta para serem disponibilizados ao público e validados pelos usuários-alvo.

Um importante ponto de atenção é a complexidade dos algoritmos de aprendizado de máquina utilizados na ferramenta. A maioria das funcionalidades são muito difíceis de executar em tempo real devido ao seu custo computacional, introduzindo limitações de interação ou exigindo *hardwares* poderosos. Pretende-se investigar ainda mais o aspecto de engenharia de aprendizado de máquina da ferramenta, com o objetivo de torná-la robusta e acessível a todos.

Mais investigações e experiências com essas técnicas e dados são necessários. A validação com usuários humanos, por exemplo, é essencial para que qualquer ferramenta de Inteligência Artificial seja disponibilizada publicamente (Sichman, 2021). Pesquisas na área de interação humano-computador estão sendo realizadas (Jaimes; Sebe, 2007), relatando resultados robustos em aplicações muito complexas. Nesse sentido, Cascella et al. (2023) avaliou o potencial e a limitação do GhatGPT em quatro cenários relacionadas à assistência médica: (1) suporte à prática clínica, (2) produção científica, (3) mau uso em medicina e pesquisa, e (4) na argumentação sobre tópicos em saúde pública. Os autores

apontam que ferramentas como o ChatGPT podem, de fato, acelerar a produção científica, pois são capazes de dar suporte a vários aspectos de uma pesquisa; como a sumarização de grandes volumes de textos médicos, prontuários e artigos científicos. Por outro lado, quando solicitado a escrever um artigo sobre um conjunto de dados falsos fornecido pelos autores, o ChatGPT criou um texto plausível e bem estruturado. Os autores realizaram diversos experimentos, cujos resultados ressaltam a importância da comunidade científica ter presente a capacidade do ChatGPT (e ferramentas similares) de gerar e disseminar informações falsas.

Os dados com os quais a ferramenta proposta lida são complexos e extensos. Temos uma preocupação real e concreta em apresentar os dados originais de forma precisa e compreensível, pois a falta de precisão nas informações apresentadas ou de interpretabilidade nos procedimentos podem ter efeitos contraproducentes, confundindo os usuários em um cenário de sobrecarga de informações.

Também devemos atentar para um aspecto importante da ferramenta proposta. Nosso objetivo neste artigo – aumentar a transparência dos processos democráticos – assume que uma ferramenta é em si abrangente, trazendo transparência e compreensibilidade às técnicas de Inteligência Artificial aplicadas (Adadi; Berrada, 2018). No entanto, a ferramenta continua sendo um conjunto de muitos módulos que as pessoas comuns não conseguem entender facilmente. Algumas direções para trabalhos futuros envolvem investigar e conduzir pesquisas sobre Inteligência Artificial explicável (Arrieta et al., 2020; Adadi; Berrada, 2018) e seu uso em aplicações envolvendo processos democráticos.

Também é importante ressaltar que, para uma verdadeira democratização das informações disponibilizadas pela ferramenta, a interface deve permitir ampla acessibilidade para pessoas com determinadas limitações, como, por exemplo, deficientes visuais. Isso pode ser um tema de atenção no futuro.

Por fim, essa ferramenta pode servir de base para outras ferramentas que possam desenvolver os outros dois eixos da democracia digital: discussão e decisão. Com a ferramenta, mostrou-se o potencial do uso de técnicas de Inteligência Artificial, aprendizagem de máquina e processamento de linguagem natural para melhorar a qualidade da informação à qual a sociedade tem acesso, o que, a longo prazo, pode aumentar a confiança na democracia.

Agradecimentos – Esta pesquisa foi parcialmente apoiada pelo Itaú Unibanco S.A., com o programa de bolsas do Programa de Bolsas Itaú (PBI), pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), Código Financeiro 001, e pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico Desenvolvimento (CNPq) (Proc. n.312360/2023-1), Brasil. Quaisquer opiniões, constatações e conclusões expressas neste manuscrito são de responsabilidade dos autores e não refletem necessariamente a visão, política oficial ou posição do Itaú-Unibanco, Capes e CNPq.

Notas

1 Disponível em: <https://chatgpt.com/>.

2 Disponível em <https://parlamento.pai.pt/>.

Referências

ADADI, A.; BERRADA, M. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE access*, IEEE, v.6, p.52138-60, 2018.

ALCOFORADO, A. et al. Zeroberto: Leveraging zero-shot text classification by topic modeling. In: PINHEIRO, V. et al. (Ed.) *Computational Processing of the Portuguese Language*. Cham: Springer International Publishing, 2022. p.125-36.

ARRIETA, A. B. et al. Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information Fusion*, v.58, p.82-115, 2020.

ASSEMBLEIA DA REPÚBLICA. 2021. Disponível em: <https://meuparlamento.pt/>. Acesso em: 29 jun. 2024.

BLEI, D. M.; NG, A. Y.; JORDAN, M. I. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, v.3, n.Jan, p.993-1022, 2003.

BONTCHEVA, K.; GORRELL, G.; WESSELS, B. Social Media and Information Overload: Survey Results. *arXiv preprint arXiv:1306.0813*, 2013.

BREINDL, Y.; FRANCQ, P. Can Web 2.0 applications save e-democracy? A study of how new internet applications may enhance citizen participation in the political process online. *International Journal of Electronic Democracy*, Inderscience Publishers, v.1, n.1, p.14-31, 2008.

CARMO, F. A. et al. Embeddings Jurídico: Representações Orientadas à Linguagem Jurídica Brasileira. In: WORKSHOP DE COMPUTAÇÃO APLICADA EM GOVERNO ELETRÔNICO (WCGE), 11, João Pessoa/PB. *Anais...* Porto Alegre: Sociedade Brasileira de Computação, p.188-99, 2023.

CASCELLA, M. et al. Evaluating the Feasibility of ChatGPT in Healthcare: An Analysis of Multiple Clinical and Research Scenarios. *Journal of medical systems*, v.47, n.33, 4 Mar. 2023.

COUNCIL OF EUROPE. *Electronic democracy ("e-democracy") – Recommendation CM/Rec(2009)1 and explanatory memorandum*. Council of Europe Publishing, 2009.

FAN, A. et al. Beyond english-centric multilingual machine translation. *Journal of Machine Learning Research*, v.22, n.107, p.1-48, 2021.

FERRAZ, T. P. et al. DEBACER: a method for slicing moderated debates. In: SBC. XVIII ENCONTRO NACIONAL DE INTELIGÊNCIA ARTIFICIAL E COMPUTACIONAL. p. 667-78, 2021.

GANDOMI, A.; HAIDER, M. Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, v.35, n.2, p.137-44, April 2015.

GROOTENDORST, M. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794*, 2022.

JAIMES, A.; SEBE, N. Multimodal human–computer interaction: A survey. *Computer Vision and Image Understanding*, v.108, n.1-2, p.116-34, 2007.

MITCHELL, R. *Web scraping with Python*: Collecting more data from the modern web. O'Reilly Media, Inc., 2018.

MOTA, C. V. 7 fatores que explicam os ataques de 8 de janeiro em Brasília. BBC News - Brasil. Disponível em <https://www.bbc.com/portuguese/articles/cye7egj6y1no>. Acesso em: 29 jun. 2024.

MULLEN, T.; MALOUF, R. A Preliminary Investigation into Sentiment Analysis of Informal Political Discourse. In: AAAI SPRING SYMPOSIUM: Computational Approaches to Analyzing Weblogs, p.159-62, 2006.

PANT, G.; SRINIVASAN, P.; MENCZER, F. Crawling the web. In: *Web Dynamics*: Adapting to Change in Content, Size, Topology and Use, Berlin, Heidelberg: Springer Berlin Heidelberg, 2004. p.153-77.

PRÉMIO TÁGIDES 2021. Disponível em: <https://www.all4integrity.org/premio-tagides/edicao2021/>. Acesso em: 29 jun. 2024.

PROCTER, R. et al. Citizen Participation and Machine Learning for a Better Democracy. *Digital Government: Research and Practice*, ACM New York, NY, USA, v.2, n.3, p.1-22, 2021.

PROTHRO, J. W.; GRIGG, C. M. Fundamental Principles of Democracy: Bases of Agreement and Disagreement. *The Journal of Politics*, Southern Political Science Association, v.22, n.2, p.276-94, 1960.

RAO, P. S. The role of english as a global language. *Research Journal of English*, v.4, n.1, p. 65-79, 2019.

ROMERA-PAREDES, B.; TORR, P. An embarrassingly simple approach to zero-shot learning. In: INTERNATIONAL CONFERENCE ON MACHINE LEARNING. p.2152-61, 2015.

SAFRANEK, R. *The emerging role of social media in political and regime change*. p.1–14. ProQuest Discovery Guides, 2012.

SICHMAN, J. S. Inteligência artificial e sociedade: avanços e riscos. *Estudos Avançados,* v.35, p.37-50, 2021.

SILVA, N. F. et al. Evaluating topic models in portuguese political comments about bills from Brazil's chamber of deputies. In: BRITTO, A.; VALDIVIA DELGADO, K. (Ed.) *Intelligent Systems*. BRACIS 2021. Lecture Notes in Computer Science, v.13074, p.104-20. Springer, Cham, 2021.

SIMON, J. et al. *Digital Democracy: The tools transforming political engagement*. [S.l.]: NESTA, UK, England and Wales 1144091, 2017. Disponível em: <https://www.nesta.org.uk/report/digital-democracy-the-tools-transforming-political-engagement/>. Acesso em: 29 jun. 2024.

TUCKER, J. A. et al. Social media, political polarization, and political disinformation: A review of the scientific literature. *SSRN Electronic Journal*, 2018.

VASWANI, A. et al. Attention is all you need. In: 31st CONFERENCE ON NEURAL INFORMATION PROCESSING SYSTEMS, Long Beach, USA, 2017.

VEDEL, T. L'idée de démocratie électronique: origines, visions, questions. In: PASCAL, P. (Ed.). *Le désenchantement démocratique*. Paris : Editions de l'Aube, 2003. p.243-66.

WATANABE, K.; ZHOU, Y. Theory-driven analysis of large corpora: Semisupervised topic classification of the UN speeches. *Social Science Computer Review,* v.40, i. 2, Apr, p.346-66, 2022.

YIN, W.; HAY, J.; ROTH, D. Benchmarking Zero-shot Text Classification: Datasets, Evaluation and Entailment Approach. In: 2019 CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING and the 9th International Joint Conference On Natural Language Processing (EMNLP-IJCNLP). p.3914-23, 2019.

YOUNG, T. et al. Recent trends in deep learning based natural language processing. *IEEE Computational Intelligence Magazine*, IEEE, v.13, n.3, p.55-75, 2018.

ZHANG, J. et al. PEGASUS: Pre-training with extracted gap-sentences for abstractive summarization. In: INTERNATIONAL CONFERENCE ON MACHINE LEARNING. p.11328-339, 2020.

RESUMO – Um dos princípios da democracia digital é informar ativamente os cidadãos e mobilizá-los para participarem no debate político. Este artigo apresenta uma ferramenta de processamento de documentos políticos públicos para tornar as informações mais acessíveis aos cidadãos e grupos profissionais específicos. Em particular, investigamos e desenvolvemos técnicas de Inteligência Artificial para mineração de textos do *Diário da Assembleia da República* de Portugal para particionar, analisar, extrair e sintetizar a informação das atas das sessões parlamentares. Desenvolvemos ainda *dashboards* que mostram as informações extraídas de forma simples e visual, como resumos de falas e tópicos discutidos. O nosso objetivo principal é, mais do que caracterizar o comportamento político, aumentar a transparência e a responsabilidade dos eleitores e das autoridades eleitas.

PALAVRAS-CHAVE: Democracia digital, Processamento de linguagem natural, Inteligência Artificial, Informação legislativa.

ABSTRACT – One of the principles of digital democracy is to actively inform citizens and mobilize them to participate in the political debate. This paper introduces a tool that processes public political documents to make information accessible to citizens and specific professional groups. In particular, we investigate and develop artificial intelligence techniques for text mining from the Portuguese *Diário da Assembleia da República* to partition, analyze, extract and synthesize information contained in the minutes of parliamentary sessions. We also developed dashboards to show the extracted information in a simple and visual way, such as summaries of speeches and topics discussed. Our main objective is to increase transparency and accountability between elected officials and voters, rather than characterizing political behavior.

KEYWORDS: Digital democracy, Natural language processing, Artificial Intelligence, Legislative information.

Alexandre Alcoforado é mestrando em Engenharia de Computação na Escola Politécnica da Universidade de São Paulo, e engenheiro elétrico pela mesma instituição.
@ – alexandre.alcoforado@usp.br / http://orcid.org/0000-0003-3184-1534.

Thomas Palmeira Ferraz é doutorando em Ciência da Computação na Télécom Paris e École Polytechnique, Institut Polytechnique de Paris. É Engenheiro pela Escola Politécnica da Universidade de São Paulo (USP) e mestre em Matemática Aplicada e Inteligência Artificial pela École Normale Supérieure Paris-Saclay (ENS).
@ – thomas.palmeira@telecom-paris.fr / http://orcid.org/0000-0002-5385-9164.

Enzo Bustos é estudante de graduação em Engenharia Elétrica com ênfase em Computação na Universidade de São Paulo. @ – enzobustos@usp.br / https://orcid.org/0000-0002-3169-4469.

André Seidel Oliveira é mestre em Engenharia de Computação pela Escola Politécnica da Universidade de São Paulo. @ – andre.seidel@usp.br / https://orcid.org/0000-0001-6551-6911.

Rodrigo Gerber é engenheiro elétrico pela Escola Politécnica da Universidade de São Paulo. @ – rodrigo.gerber@usp.br / http://orcid.org/0000-0002-8120-7972.

Gian Lucca du Mont Santoro é estudante de graduação em Matemática Aplicada e Computacional (com qualificação em Estatística Econômica) no Instituto de Matemática e Estatística da Universidade de São Paulo, e cientista de dados no Itaú Unibanco.
@ – gian.santoro@usp.br / http://orcid.org/0000-0001-6731-7940.

Israel Campos Fama é doutorando em Engenharia de Computação na Escola Politécnica da Universidade de São Paulo. É engenheiro mecânico pela Universidade Federal do Ceará (UFC) e mestre em Engenharia Aeronáutica pelo Instituto Tecnológico de Aeronáutica (ITA). @ - israelfama@usp.br / http://orcid.org/0000-0001-6325-4153.

Bruno Veloso é professor auxiliar da Faculdade de Economia da Universidade do Porto. É pesquisador sênior da Liaad Inesc Tec na área de Machine Learning, Data Streams e AutoML, e membro da associação espanhola de inteligência artificial. Suas áreas de interesse incluem sistemas multiagentes, agentes de software inteligentes, mineração de dados e fluxos de dados. @ – bveloso@fep.up.pt / http://orcid.org/0000-0001-7980-0972.

Fábio Levy Siqueira é professor doutor no Departamento de Engenharia de Computação e Sistemas Digitais da Escola Politécnica da Universidade de São Paulo. É pesquisador na área de Engenharia de Software, com interesses de pesquisa em Engenharia de Requisitos, Métodos Ágeis e Engenharia Dirigida por Modelos. @ – levy.siqueira@usp.br / http://orcid.org/0000-0002-7550-2000.

Anna Helena Reali Costa é professora titular de Engenharia de Computação na Universidade de São Paulo (USP). Doutora pela USP, pesquisadora convidada no Karlsruhe Institute of Technology, Alemanha, e na Carnegie Mellon University, EUA. É diretora do Centro de Ciência de Dados (C2D), uma parceria entre a USP e o Banco Itaú-Unibanco, e membro do Centro de Inteligência Artificial USP-Fapesp-IBM (C4AI).
@ – anna.reali@usp.br / http://orcid.org/0000-0001-7309-4528.

Recebido em 26.1.2023 e aceito em 16.11.2023.

[I,III,IV,V,VII,IX,X] Universidade de São Paulo, Escola Politécnica, São Paulo, Brasil.

[II] Telécom Paris, Institut Polytechnique de Paris, Palaiseau, França.

[VI] Universidade de São Paulo, Instituto de Matemática e Estatística, São Paulo, Brasil.

[VIII] Universidade do Porto e INESC TEC, Porto, Portugal.