

Grafos de conhecimento - explorando o potencial de modelos de geração de linguagem

Marcio Carneiro dos Santos

Resumo:

Apresenta-se estudo exploratório sobre o potencial de utilização de modelos de geração de linguagem como ferramentas auxiliares no trabalho do jornalismo guiado por dados e das narrativas automatizadas, através da construção de grafos de conhecimento, para facilitar a visualização de relações entre entes detectados em corpus textuais, bem como a possibilidade de usar tais modelos de forma mais assertiva, a partir de bases de conhecimento específicas, de acordo com a necessidade dos usuários. A metodologia utilizada foi o desenvolvimento de prova de conceito, usando o modelo GPT da *OpenAI*, bem como sua API (*Application Programming Interface*). Os experimentos demonstraram potencial de utilização e a capacidade de serem ampliados para operar em contextos de maior escala.

Palavras-chave: Inteligência Artificial. Processamento de Linguagem Natural. Jornalismo Guiado por Dados.

Knowledge graphs - exploring the potential of language generation models

Abstract:

An exploratory study is presented on the potential of using language generation models as auxiliary tools in the work of data-driven journalism and automated narratives, through the construction of knowledge graphs, to facilitate the visualization of relationships between detected entities in textual corpus, as well as the possibility of using such models in a more assertive way, based on specific knowledge bases, according to the users' needs. The methodology used was the development of proof of concept, using OPENAI's GPT model, as well as its API (*Application Programming Interface*). The experiments demonstrated potential for use and the ability to be expanded to operate in larger-scale contexts.

Keywords: Artificial intelligence. Natural Language Processing. Data-Driven Journalism.

Recebido em: 17.06.23
Aprovado em: 10.12.23

Marcio Carneiro dos Santos

Doutor em Tecnologias da Inteligência e Design Digital/PUC-SP. Professor Adjunto do Departamento de Comunicação Social/Universidade Federal do Maranhão (UFMA) na área de Jornalismo em Redes Digitais.

E-mail: mcszen@gmail.com

Estudos em Jornalismo e Mídia
v. 20, n. 2, jul./dez. 2023.
ISSNe 1984-6924

Introdução

A linguagem natural é a forma mais comum de comunicação entre os seres humanos. Ela é caracterizada pela sua complexidade, ambiguidade e variabilidade. A compreensão e o processamento automático dessa linguagem são desafios significativos para a área de Inteligência Artificial (IA).

O Processamento de Linguagem Natural (PLN) é uma área da Ciência da Computação que se insere entre as diversas possibilidades de utilização da IA. Especificamente, o PLN visa permitir que computadores entendam, interpretem e processem a linguagem humana de maneira eficaz. É um campo interdisciplinar que combina conhecimentos de linguística, inteligência artificial e ciência da computação para desenvolver técnicas e algoritmos capazes de lidar com os principais aspectos da comunicação humana.

O objetivo central do PLN é capacitar os computadores a compreender, interpretar e produzir texto em linguagem natural. Isso envolve diversas tarefas, como: reconhecimento de fala convertendo áudio em texto; análise morfológica identificando a estrutura das palavras, como raízes, sufixos e prefixos; análise sintática analisando a estrutura gramatical de uma sentença; análise semântica compreendendo o significado das palavras e frases em um determinado contexto; desambiguação lidando com a ambiguidade presente na linguagem natural e, por fim, geração de texto produzindo textos coerente e naturais em linguagem humana.

Para operar com a maioria dessas funções o PLN utiliza modelos de linguagem natural. Um modelo de linguagem natural é uma representação estatística e probabilística das estruturas e padrões da linguagem humana. É uma forma de modelar a probabilidade de ocorrência de sequências de palavras ou caracteres em um determinado contexto. Numa comparação rudimentar, seria como a função de autocompletar em determinados aplicativos, como o WhatsApp, na qual uma palavra é sugerida após cada etapa da digitação do usuário, só que de uma forma muito mais elaborada e complexa.

Os modelos de linguagem natural são treinados em grandes quantidades de texto para aprender as probabilidades de ocorrência de palavras ou sequências de palavras. Essas probabilidades são usadas para prever a próxima palavra em uma sequência, avaliar a fluidez de um texto ou gerar texto de maneira coerente.

GPT é um modelo de linguagem pré-treinado (*Pre-trained Language Model*) desenvolvido pela empresa *OpenAI*. Ele foi desenhado usando uma grande quantidade de documentos da internet e é capaz de gerar texto natural e coerente, quando recebe um *input*, também conhecido como *prompt*, com uma determinada tarefa ou contexto. O modelo GPT pode ser usado em várias aplicações, incluindo *chatbots*, geradores de conteúdo e sistemas de tradução automática.

O produto CHAT GPT, lançado publicamente em novembro de 2022, é uma implementação do modelo GPT em formato de robô conversacional; um *chatbot* que após cinco dias de exposição pública conseguiu cerca de um milhão de usuários (*OpenAI*, 2023), quebrando todos os recordes conhecidos de adoção tecnológica e angariando enorme atenção da mídia e da sociedade em geral.

Este e outros produtos recentes como *Midjourney*, *Stable Diffusion* e *Dall-E* para citar apenas alguns, inserem-se na subárea da inteligência artificial denominada de *generative AI*, ou seja, inteligência artificial gerativa ou generativa (IAG) que, de forma simples, busca desenvolver soluções orientadas a simular a própria criatividade humana, transformando entradas de texto em novos textos ou até imagens, todos construídos a partir de enormes bases de dados.

Apesar da atenção recente, é bom ressaltar que a inteligência artificial em suas áreas de origem tem décadas de desenvolvimento e já é utilizada em diversos contextos, como na detecção de *spams* nas caixas de e-mail (algoritmos de

classificação) e em plataformas nas quais o histórico de utilização ou navegação serve de base para um contato mais próximo com o utilizador, como em sugestões baseadas em compras anteriores nos *marketplaces* ou filmes em plataformas de *streaming* (algoritmos de recomendação).

Especificamente, modelos de linguagem natural estão na base de várias aplicações e produtos que conhecemos, tais como assistentes virtuais em plataformas como *Siri*, *Alexa* e *Google Assistant* que utilizam modelos para entender comandos de voz, responder a perguntas e realizar tarefas, como agendar compromissos, reproduzir músicas e fornecer informações em linguagem natural.

Outra funcionalidade bastante comum é a tradução automática, através de serviços como o *Google Tradutor*, que utiliza processamento de linguagem natural para analisar e compreender o texto em um idioma e gerar uma tradução coerente em outro idioma.

Com larga utilização no marketing digital e em diversos tipos de pesquisas, algoritmos com base no PLN são empregados para analisar a opinião e o sentimento expressos em textos, como comentários em mídias sociais ou avaliações de produtos, permitindo que as empresas compreendam a percepção do público em relação a determinados temas, a partir de uma análise totalmente automatizada de sentimentos ou de forma híbrida, integrando algoritmos e curadoria humana, como propõe Santos (2018).

Chatbots e atendimento ao cliente são também aplicações bem comuns operadas a partir de modelos de linguagem natural. Nessas soluções, por exemplo, empresas fornecem suporte e interagem com os clientes de maneira automatizada. Esses *chatbots* podem responder a perguntas, auxiliar na resolução de problemas e realizar transações básicas.

Na finalidade de sumarização automática, o PLN pode ser aplicado para resumir grandes volumes de texto, como notícias, artigos científicos ou documentos legais, fornecendo uma visão geral das informações de maneira concisa.

Por fim, para pesquisa na web e recuperação de informações, os mecanismos de busca utilizam técnicas de PLN para compreender a intenção do usuário e fornecer resultados relevantes para suas consultas em linguagem natural.

As conexões possíveis entre inteligência artificial e jornalismo têm origens nas possibilidades de aplicação de abordagens quantitativas no processo de produção de conteúdo informativo, como demonstrado em Dos Santos (2015), bem como em todo um conjunto já consolidado de estudos ligados à interdisciplinaridade entre os campos da Tecnologia e da Comunicação (Júnior, Dos Santos, 2013), bem como o tema das narrativas automatizadas já extensamente explorado em textos anteriores, como em Carreira (2017), que fala de um jornalismo feito por não humanos, e Višnovský et al. (2019), que propõem um robojornalismo, resultado da interconexão entre informática, estatística e reportagem. Outros textos usam termos como jornalismo automatizado e jornalismo robótico (Maier, 2002; Dalen, 2012; Carlson, 2014; Clerwall, 2014; Lewis & Usher, 2014; Latar, 2015; Santos, 2014; Santos, 2016a; Graefe, 2016) ao tratarem de processos semelhantes.

IA e Jornalismo

No campo dos estudos do jornalismo, o interesse por narrativas automatizadas tem sido crescente, mesmo quando os casos estudados tratam de soluções que não são necessariamente desenvolvidas usando inteligência artificial, mas sim pelo que conhecemos como automatização por regras ou até, de forma mais simples, conteúdo dinâmico. Especificamente sobre a relação entre inteligência artificial (IA) e jornalismo vários recortes têm sido abordados por pesquisadores nacionais e internacionais.

Com a ascensão das plataformas digitais e o aumento da produção e consumo de conteúdo on-line, o papel da IA no jornalismo tem se tornado cada vez mais relevante. Nesse contexto, surgem desafios e oportunidades para os profissionais da área, que devem entender e se adaptar às mudanças.

Um dos principais desafios é a possibilidade de a IA substituir o trabalho humano, o que levanta a questão se a IA irá suplantar a nossa inteligência. Segundo Kaufman (2019), a IA é uma tecnologia de propósito geral, assim como a eletricidade, que pode ser aplicada em diversas áreas. Embora a IA possa ajudar a automatizar tarefas repetitivas e aprimorar a produção de conteúdo, ela não pode entregar conteúdo que dispense revisão ou curadoria posterior. Afinal, a IA é alimentada por dados e algoritmos, e não tem capacidade de criatividade e de interpretação.

Outro desafio que o jornalismo enfrenta é a reestruturação do setor causada pela ascensão das plataformas digitais, como *Facebook*, *Google* e *Twitter*. Bell *et al.* (2017) destacam que as plataformas sociais têm influenciado o jornalismo em si, incentivando a produção de conteúdo para engajamento do público em vez de focar na produção de notícias com valor cívico. A IA pode ser usada para produzir conteúdo mais personalizado para o público, o que pode aumentar o engajamento, mas ao mesmo tempo pode reforçar bolhas de informação e desinformação.

Apesar dos desafios, a IA também apresenta oportunidades para o jornalismo. Uma das principais vantagens é a possibilidade de coletar e analisar grandes quantidades de dados em tempo real, permitindo aos jornalistas identificar tendências e padrões, e produzir conteúdo mais preciso e relevante. A IA também pode ser usada para automatizar tarefas repetitivas, como a geração de relatórios e análises estatísticas.

Dos Santos (2020) propõe algumas possibilidades adicionais para o uso de IA nas redações, incluindo análise preditiva de engajamento futuro para notícias, organização e planejamento para alocação de recursos em coberturas, bem como o que chama de personalização estendida, oportunizando um acréscimo na percepção de valor do consumidor da informação para o conteúdo ofertado, com a geração de versões específicas, produzidas a partir de padrões de consumo anterior deste usuário dentro da plataforma de notícias.

Além disso, a IA pode ser usada para criar novas formas de narrativas jornalísticas (Santos, 2016). Bertocchi (2016) destaca que a narrativa é a maneira como o jornalista organiza eventos múltiplos em uma única história, e a IA pode ser usada para criar narrativas mais envolventes e interativas. Por exemplo, a IA pode ser usada para criar experiências de realidade virtual imersivas ou para gerar conteúdo em diferentes formatos, como infográficos e vídeos.

Outra vantagem da IA no jornalismo é a possibilidade de produzir conteúdo em larga escala e em diferentes idiomas. Isso pode ajudar a aumentar a diversidade e a acessibilidade do conteúdo jornalístico, permitindo que ele chegue a um público mais amplo. Pellanda *et al.* (2017) destacam que a IA também pode ser usada para entender hábitos e circunstâncias contextuais em que as informações podem ser reunidas, até mesmo antes do usuário solicitar.

Outra oportunidade oferecida pela IA no jornalismo é o uso de algoritmos para a produção de notícias. Linden (2018) destaca que a produção de notícias por meio de algoritmos, conhecida como jornalismo algorítmico, computacional ou robótico, já é uma realidade e pode ajudar a agilizar a produção e distribuição de conteúdo. No entanto, é importante lembrar que a produção de notícias por meio de algoritmos pode gerar questões éticas, como a verificação da fonte e a imparcialidade do conteúdo.

Os algoritmos de IA também têm sido usados para combater a desinformação e as *fake news*, que são um grande problema no ambiente digital. Kaufman e Santaella (2020) destacam que o controle dos algoritmos de IA é uma preocupação central do jornalismo, pois eles podem ser usados também para difundir desinformação. Nesse sentido, a IA pode ser usada para identificar e verificar informações falsas e gerar conteúdo mais confiável e preciso.

Além dos pesquisadores nacionais, o debate sobre o tema também ocorre, inclusive há mais tempo, em muitos países.

Um estudo realizado por Newman (2018) revelou que quase três quartos dos profissionais de mídia já usam algum tipo de inteligência artificial em seus trabalhos. No entanto, o uso da IA no jornalismo ainda está em sua infância, e muitas questões éticas, legais e práticas precisam ser abordadas.

Um dos principais desafios é a garantia da qualidade e da precisão das informações produzidas pela IA. Guzman e Lewis (2020) argumentam que é necessário desenvolver uma agenda de pesquisa para a comunicação humano-máquina, a fim de explorar questões fundamentais, como a compreensão e a confiança na IA e as formas de integrar a *expertise* humana com a IA.

Lewis, Guzman e Schmidt (2019) destacam que a IA é frequentemente vista como uma fonte de informação, em vez de um canal de comunicação. Isso pode levar a uma falta de transparência sobre como a IA toma decisões e quais critérios são usados para determinar quais informações são relevantes. Essa falta de transparência pode prejudicar a confiança do público nas notícias produzidas pela IA.

A cobertura da IA na mídia também é um tema importante. Brennen (2018) analisou a cobertura da IA em seis meios de comunicação no Reino Unido e descobriu que a cobertura era frequentemente sensacionalista e focada em eventos recentes, em vez de fornecer uma análise crítica das implicações mais amplas da IA.

Hansen, Roca-Sales, Keegan e King (2017) destacam que a IA também pode ser usada para identificar tendências e padrões em grandes conjuntos de dados, o que pode ajudar os jornalistas a encontrar novas histórias e a investigar questões complexas. Stray (2019) aponta que a IA também pode ajudar a acelerar o processo de investigação jornalística, permitindo que os jornalistas analisem grandes quantidades de dados em menos tempo.

No entanto, a automação completa do jornalismo também tem sido criticada por muitos estudiosos. Broussard e Diakopoulos (2019) argumentam que a IA não é capaz de realizar julgamentos éticos e não pode substituir completamente a *expertise* e a intuição humana. Galily (2018) destaca que a automação pode levar à perda de empregos para os jornalistas e pode reduzir a diversidade de opiniões e perspectivas na mídia.

Como apontam Ali e Hassoun (2019), a IA tem o potencial de ajudar os jornalistas a superar muitos dos desafios enfrentados atualmente na indústria, como a sobrecarga de informações e a falta de tempo para análise e investigação. No entanto, a implementação da IA no jornalismo deve ser cuidadosamente planejada e monitorada, a fim de garantir que os benefícios sejam maximizados e os riscos minimizados.

Neste artigo, operamos com o recorte mais específico do jornalismo investigativo e o potencial uso de soluções de IA generativa (IAG) de última geração para servir de ferramenta, basicamente em dois cenários:

- a) a geração de visualizações de redes de relacionamento, conhecidas como grafos de conhecimento (*knowledge graphs*);
- b) a utilização de bases de documentos menores, contudo mais ligadas à temática de utilização a partir da qual se deseja operar, de forma a gerar resultados mais assertivos.

Objetivos

O objetivo geral deste trabalho é descrever um experimento utilizando a nova geração de IAG para apoiar o trabalho de jornalistas investigativos.

Especificamente o que nos moveu foram os propósitos de:

- a) desenvolver novas formas de utilização da IAG para gerar visualizações de dados no formato de grafos de conhecimento;
- b) utilizar o modelo de geração de textos GPT num ambiente focado em um

conjunto menor de documentos do que sua base de treinamento original, a fim de conseguir mais assertividade nos resultados que ele nos retorna.

Métodos

O processamento de linguagem natural - PLN (em inglês, NLP - *Natural Language Processing*) é uma subárea da inteligência artificial que se concentra no desenvolvimento de algoritmos capazes de processar, analisar e compreender a linguagem natural utilizada pelos seres humanos. Isso significa que, ao contrário das máquinas que são programadas para seguir comandos específicos, o PLN permite que as máquinas processem dados não estruturados, como textos, de maneira a entender a semântica e a sintaxe desses dados.

As possibilidades de análise de texto utilizando PLN são muitas. Dentro desse campo, uma delas é conhecida como NER. O Reconhecimento de Entidades Nomeadas (NER, do inglês *Named Entity Recognition*) é uma das técnicas mais utilizadas no Processamento de Linguagem Natural (Grus, 2016). Essa técnica tem como objetivo identificar e classificar elementos específicos em um texto, como nomes de pessoas, organizações, locais, datas, valores monetários e outros tipos de entidades nomeadas.

O NER utiliza algoritmos de aprendizado de máquina para analisar o texto e identificar os padrões linguísticos que indicam a presença de entidades nomeadas. Esses algoritmos utilizam técnicas de reconhecimento de padrões e de análise sintática para identificar as entidades presentes no texto.

A análise de NER pode ser utilizada para diversas finalidades, tais como:

Análise de dados: O NER pode ser utilizado para extrair informações relevantes de grandes volumes de dados não estruturados. Por exemplo, em uma empresa de seguros, pode-se usar o NER para identificar nomes de clientes, datas de vencimento de apólices e valores de prêmios em contratos de seguros;

Monitoramento de mídias sociais: Empresas e organizações podem usar o NER para monitorar a presença de sua marca em redes sociais, identificando menções positivas ou negativas, opiniões dos clientes, entre outros;

Análise de texto em tempo real: O NER pode ser utilizado em aplicações que exigem análise de texto em tempo real, como *chatbots*, assistentes virtuais, sistemas de análise de sentimento, entre outros;

Análise de documentos jurídicos: O NER pode ser utilizado para identificar nomes de partes, datas e valores em documentos jurídicos, acelerando processos legais;

Análise de notícias: O NER pode ser utilizado para identificar as entidades nomeadas em notícias, permitindo que os jornalistas identifiquem rapidamente as pessoas, organizações e locais mencionados na história.

As técnicas de Processamento de Linguagem Natural (PLN) podem ser combinadas com visualizações de dados para proporcionar insights e compreensão mais intuitiva das informações extraídas do texto. Na Tabela 1, estão algumas maneiras de utilizar visualizações de dados a partir das técnicas de PLN:

Tabela 1 – Lista de possibilidades e formatos de visualização de dados que podem ser utilizados a partir de técnicas de processamento de linguagem natural.

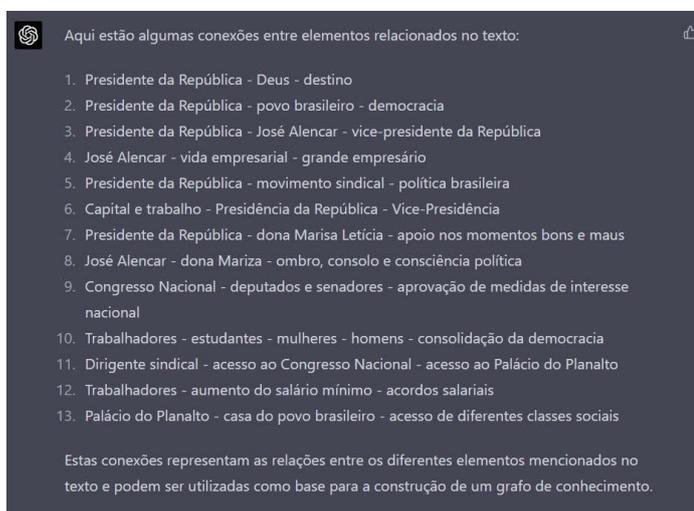
| Tipo de Visualização | Explicação |
|----------------------|--|
| <i>Word Cloud</i> | Representação visual das palavras mais frequentes em um texto. É útil para identificar os termos mais relevantes e dominantes. |

| | |
|----------------------------|--|
| Histograma de Frequência | Visualização da distribuição de frequência das palavras em um <i>corpus</i> de texto. Ajuda a identificar os termos mais comuns e sua frequência relativa. |
| Gráficos de Redes | Visualização das interações entre palavras ou entidades extraídas do texto. Permite identificar <i>clusters</i> de termos relacionados e suas interações. |
| Visualização de Sentimento | Representação gráfica da distribuição de sentimentos (positivo, negativo, neutro) em um conjunto de textos. Auxilia na compreensão do sentimento predominante. |
| Análise de Tópicos | Visualização da distribuição de tópicos em um <i>corpus</i> de texto. Permite identificar a prevalência de diferentes temas e áreas de discussão. |
| Mapas de Calor | Representação visual da coocorrência de palavras ou tópicos em um conjunto de documentos. Destaca as relações e padrões de associação entre termos. |
| Gráficos Temporais | Visualização da evolução de termos, tópicos ou sentimentos ao longo do tempo. Permite identificar tendências e padrões em dados textuais sequenciais. |
| Diagramas de Venn | Utilizados para visualizar a sobreposição ou interseção de conjuntos de palavras, temas ou entidades. Auxilia na análise de relações entre conjuntos de dados. |

Fonte: do autor.

Em nosso experimento escolhemos o trecho de um texto entre os discursos presidenciais de posse, coletados no próprio site da Presidência da República (Fig. 1). O texto foi o do presidente Lula em 2007, a partir do qual solicitamos para o CHAT GPT realizar um reconhecimento de entidades e, também, a identificação das relações entre elas que ele conseguisse identificar.

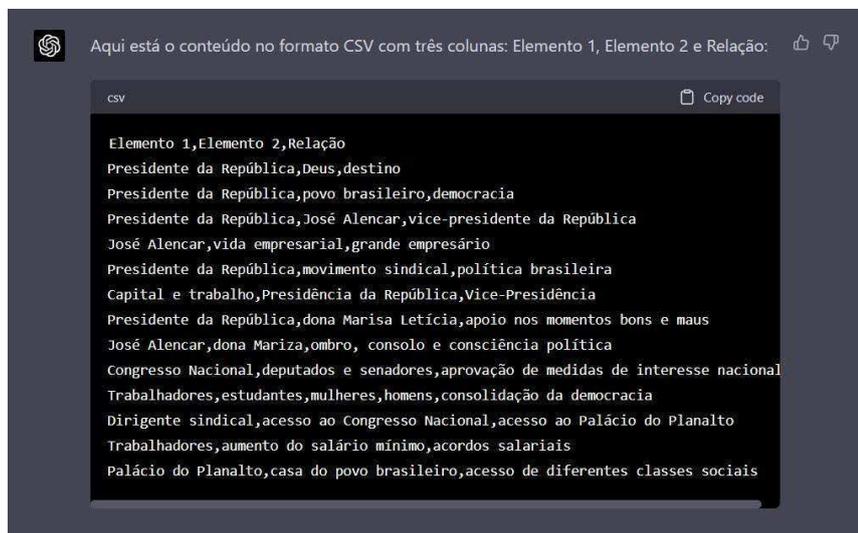
Figura 1 – Inserção do trecho do discurso do presidente Lula 2007 no prompt do CHAT GPT solicitando a ele que realizasse um reconhecimento de entidades.



Fonte: do autor.

Dentro dos procedimentos do método que desenvolvemos, a lista tabular que incluía as duas entidades em colunas diferentes e a relação entre elas numa terceira coluna foi transformada, para a extensão CSV, formato estruturado em arquivo de texto que consegue, de forma simples, representar a tabela original e depois ser exportado e utilizado em outras aplicações (Fig. 2).

Figura 2 – Conversão do resultado para o formato CSV de forma a permitir a exportação e uso no software de geração de grafos Gephi.



```

csv
Elemento 1,Elemento 2,Relação
Presidente da República,Deus,destino
Presidente da República,povo brasileiro,democracia
Presidente da República,José Alencar,vice-presidente da República
José Alencar,vida empresarial,grande empresário
Presidente da República,movimento sindical,política brasileira
Capital e trabalho,Presidência da República,Vice-Presidência
Presidente da República,dona Marisa Letícia,apoio nos momentos bons e maus
José Alencar,dona Mariza,ombro, consolo e consciência política
Congresso Nacional,deputados e senadores,aprovação de medidas de interesse nacional
Trabalhadores,estudantes,mulheres,homens,consolidação da democracia
Dirigente sindical,acesso ao Congresso Nacional,acesso ao Palácio do Planalto
Trabalhadores,aumento do salário mínimo,acordos salariais
Palácio do Planalto,casa do povo brasileiro,acesso de diferentes classes sociais

```

Fonte: do autor.

O objetivo final é, através das etapas que detalharemos a seguir, usar esse resultado para gerar um grafo de conhecimento.

Um grafo de conhecimento (ou *knowledge graph*) é uma estrutura de dados que representa conhecimento em um formato que é fácil de entender tanto por humanos quanto por máquinas. Essa estrutura de dados é composta por entidades (como pessoas, lugares, coisas, conceitos) e relações entre elas.

O grafo de conhecimento é construído a partir de informações extraídas de diversas fontes, como bancos de dados, textos, imagens, vídeos, etc. As informações são organizadas em entidades e suas relações e, em seguida, são representadas em uma estrutura de grafo, onde os nós representam as entidades e as arestas representam as relações.

O grafo de conhecimento é usado em diversas áreas, como inteligência artificial, processamento de linguagem natural, ciência da computação e *web* semântica. Ele é usado para ajudar as máquinas a entender melhor o significado do texto, para melhorar os resultados de busca e para ajudar a criar sistemas de recomendação mais precisos. Ele também é usado em aplicações de *big data* e análise de dados, onde ajuda a identificar padrões e insights úteis.

Com o experimento, nosso intuito é demonstrar que, de forma relativamente rápida, a partir de uma massa de textos, o jornalista investigativo poderá construir uma visualização que lhe permitirá não só entender, mas também informar para os outros as relações estabelecidas pelos entes que aparecem neste material.

Resultados

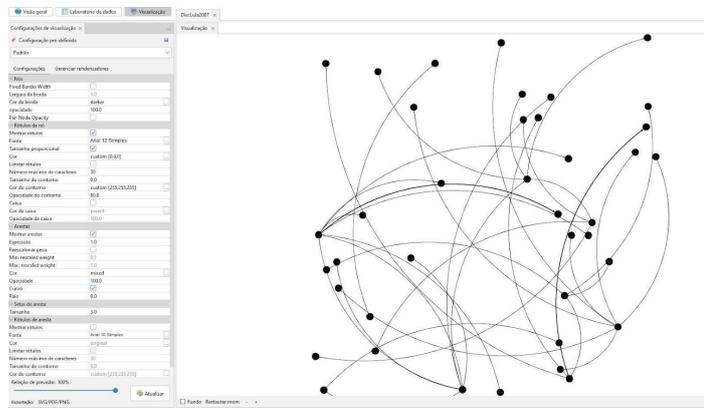
Por fim, com o arquivo em mãos, levamos os dados ao *Gephi*, software que consegue gerar, a partir de dados estruturados, como nosso arquivo CSV, uma representação visual da rede formada pelas entidades e a relação que estabelecem dentro do texto analisado.

As limitações em termos de tamanho deste arquivo não nos permite detalhar o processamento interno feito no software *Gephi*. Contudo, este software é bastante docu-

mentado. Na internet e no próprio site do desenvolvedor existe grande material informativo e tutoriais para executar essa etapa. Em termos simples, o que fizemos foi importar nosso arquivo CSV, contendo as relações encontradas no texto pelo modelo de linguagem, para que o *Gephi*, a partir dos dados, montasse um grafo, ou seja, a visualização gráfica da rede formada pelas entidades extraídas via NER e as relações que identificou.

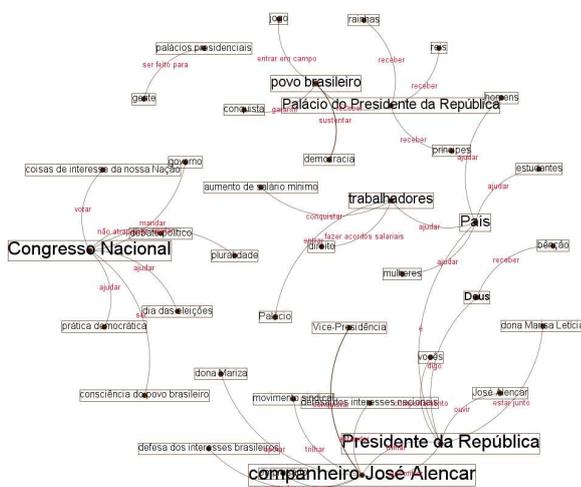
Na imagem abaixo (Fig. 3) é fácil perceber como após o tratamento no *Gephi*, o conjunto de relações extraídas do texto começa a se materializar visualmente. Cada ponto representa uma entidade detectada e cada linha curva representa uma relação entre duas entidades que o CHAT GPT apontou na sua análise NER. Já a Figura 4 mostra o resultado final no *Gephi* com a geração do grafo que demonstra com apreensão mais imediata as relações de entidades contidas no discurso. As legendas dos pontos, colocadas ao lado de cada um deles, tem corpo proporcional à intensidade ou a frequência da presença dos mesmos na rede.

Figura 3 – Fase inicial de tratamento no software Gephi a partir do arquivo CSV de forma a permitir a construção de visualização das entidades e das relações detectadas no formato de um grafo, ou seja, uma representação gráfica desta rede.



Fonte: do autor.

Figura 4 – Resultado final no Gephi com a geração do grafo que demonstra com apreensão mais imediata as relações de entidades contidas no discurso.



Fonte: do autor.

⁶ No original: “It refers simultaneously to the appearance of witnesses in media reports, the possibility of media themselves bearing witness, and the positioning of media audiences as witnesses to depicted events” (FROSH; PIN-CHEVSKI, 2009).

Conclusões

O experimento através da prova de conceito desenvolvida usando as funcionalidades de análise do CHAT GPT, ora apresentado, tem o intuito apenas de demonstrar um leque possível e, talvez, ainda desconhecido, de possibilidades para utilização da ferramenta na área do jornalismo.

É bem comum que o profissional se depare com conjuntos de textos e precise de uma estratégia inicial para começar a apreender as informações, padrões e histórias que, efetivamente, aquele determinado *corpus* contém.

É o que na Ciência de Dados chamamos de EDA - do inglês *exploratory data analysis* - processo inicial que não responderá diretamente as questões específicas que queremos perguntar ou extrair de determinado conjunto de informações, mas pode nos dar uma ideia inicial de com que estamos lidando e quais suas características ou padrões básicos.

A proposta da construção de gráficos de conhecimento neste contexto nos parece útil com esse propósito de primeiro contato, semelhante ao papel que as principais métricas da estatística descritiva nos proporcionam.

A partir da prova de conceito apresentada, espera-se que novas utilizações e apropriações destas ferramentas sejam desenvolvidas, lembrando que, pelo caráter de experimento apenas, o trecho do texto usado foi pequeno. Contudo, a partir do uso da API do próprio modelo GPT, agora em sua versão 4, tal procedimento pode ser escalado e levado a conjuntos maiores e mais próximos do que um jornalista investigativo encontraria no mundo real.

Em última análise, entendemos que a IA não é uma ameaça ao jornalismo, mas sim uma ferramenta poderosa que pode ajudar a aprimorar a qualidade das informações produzidas e a democratizar o acesso à informação.

É importante que os profissionais de mídia continuem a se envolver em discussões críticas sobre a IA e a explorar maneiras de integrar a tecnologia de forma responsável e eficaz no campo do jornalismo. Esperamos que o experimento aqui apresentado contribua com este processo.

Referências bibliográficas

BERTOCCHI, D. **Dos dados aos formatos: a construção de narrativas no jornalismo digital**. São Paulo: Summus, 2016.

BELL, E; OWEN, T; BROWN, P; HAUKA, C. **A Imprensa Nas Plataformas: Como O Vale Silício Reestruturou o Jornalismo**. Columbia University Academic Commons, 2017.

BRENEN, J. **An industry-led debate: How UK media cover artificial intelligence**. Reuters Institute for the Study of Journalism, 2018.

BROUSSARD, M; DIAKOPOULOS, N. Artificial intelligence and journalism. **Journalism & Mass Communication Quarterly**, v. 96, n. 3, p. 745-759, 2019.

CARLSON, M. The Robotic Reporter: Automated journalism and the redefinition of labor, compositional forms and journalistic authority. In: LEWIS, Seth C. (org.). **Digital Journalism**. Vol. 3, nº 3. New York: Taylor&Francis Online, 2014.

CARREIRA, K. **Notícias Automatizadas** – A evolução que levou o jornalismo a ser feito por não humanos. Universidade Metodista, 2017. Disponível em <http://tede.metodista.br/jspui/bitstream/tede/1671/1/KrishmaCouraCarreira.pdf>. Acessado em 11/04/2019.

CLERWALL, C. Enter the Robot Journalist: User's perception of automated content. In: **Journalism Practice**. Special Issue – Future of Journalism in an age of digital media and economic uncertainty. Volume 8, Issue 5. New York: Taylor & Francis Online, 2014.

CONCEIÇÃO, V; CHAGAS, A. O pesquisador e a divulgação científica em contexto de cibercultura e inteligência artificial. **Acta Scientiarum. Education**, v. 42, e46232, 2020.

DALEN, A. The Algorithms Behind the Headlines: How machine-written news redefines the core skills of human journalists. In: **Journalism Practice**. Volume 6, Issue 5-6. New York: Routledge, 2012.

DOS SANTOS, M. Métodos digitais e a memória acessada por APIs: desenvolvimento de ferramenta para extração de dados de portais jornalísticos a partir da WayBack Machine. **Revista Observatório**, v. 1, n. 2, p. 207-228, 2015.

DOS SANTOS, M. Data-Driven Journalistic Operation: Reshaping the Idea of News Values with Algorithms, Artificial Intelligence and Increased Personalization. **Brazilian journalism research**, v. 16, n. 3, p. 458-475, 2020.

GALILY, Y. Artificial intelligence and sports journalism: Is it a sweeping change? **Technology in Society**, v. 54, p. 28-34, 2018.

GRAEFE, A. Guide to Automated Journalism. Tow Center for **Digital Journalism**. Janeiro, 2016. Disponível em <https://www.gitbook.com/book/towcenter/guide-to-automated-journalism/details>. Acessado em 20/01/2016.

GRUS, J. **Data Science do Zero**. Rio de Janeiro: Alta Books, 2016.

GUZMAN, A.; LEWIS, S; SCHMIDT, T. Automation, journalism, and human-machine communication: Rethinking roles and relationships of humans and machines in news. **Digital Journalism**, v. 7, n. 6, p. 797-813, 2019.

HANSEN, M; ROCA-SALES, M; KEEGAN, J; KING, G. Artificial intelligence: Practice and implications for journalism. **Columbia Journalism School**, 2017.

JÚNIOR, J; DOS SANTOS, M. **Comunicação, tecnologia e inovação: estudos interdisciplinares de um campo em expansão**. Buqui Livros Digitais, 2013.

KAUFMAN, D. **O papel dos algoritmos de inteligência artificial nas redes sociais**. Revista Famecos, v. 27, n. 1, p. 1-16, 2020.

KAUFMAN, D. **A inteligência artificial irá suplantar a inteligência humana?** São Paulo: Alta Books, 2019.

LATAR, N. The Robot Journalism in the Age of Social Physics: **The end of human journalism?** In: The New World of Transitioned Media. Springer, 2015.

LATAR, N. **Robot journalism: Can human journalism survive?** Social DNA, 2018.

LEWIS, S; GUZMAN, A; SCHMIDT, T. Artificial intelligence and communication: A Human–Machine Communication research agenda. **New Media & Society**, v. 22, n. 8, p. 1368-1387, 2020.

LEWIS, S; USHER, N. Code. Collaboration and The Future of Journalism: A case study of the Hacks/Hackers global network. In: **Digital Journalism**. Routledge Online, 2014.

LINDEN, C. **Algoritmos para Jornalismo**: o futuro da produção de notícias. **Líbero**, v. 21, n. 42, p. 153-163, 2018.

MAIER, S. Numbers in the News: A Mathematics Audit of a Daily Newspaper. **Journalism Studies** 3. New York: Taylor & Francis Online, 2002.

OPENAI - Evento Development Day [Online] - Novembro, 2023.

PELLANDA, E; PASE, A; NUNES, A. **Mobilidade e jornalismo digital contemporâneo**: fases do jornalismo móvel ubíquo e suas características. **Jornalismo Móvel em perspectiva**, 2017.

PRADO, M. **Fake News e Inteligência Artificial**: O poder dos algoritmos na guerra da desinformação. São Paulo: Novatec Editora, 2022.

SANTOS, M. Textos Gerados por Software – Surge Um Novo Gênero Jornalístico? **Anais do XXXVII Congresso Brasileiro de Ciências da Comunicação**. Foz do Iguaçu, 2014. Disponível em <http://www.intercom.org.br/papers/nacionais/2014/resumos/R9-2066-1.pdf>. Acesso em 11/04/2019.

SANTOS, M. Narrativas Automatizadas e a Geração de Textos Jornalísticos: A estrutura de organização do lead traduzida em código. **Brazilian Journalism Research**, v. 12, n. 2, p. 258-276, 2016.

SANTOS, M. Internet das Coisas e Sistemas Inteligentes no Jornalismo. **Comunicação e Inovação**. Vol 17, Págs 21-39, 2016. Disponível em: http://seer.uscs.edu.br/index.php/revista_comunicacao_inovacao/article/view/3769. Acesso em 27/03/2020.

SANTOS, M. Inteligência híbrida e análise de sentimentos: integrando curadoria humana e coleta de dados automatizada para avaliar a comunicação de governo. **Revista Conexão - Comunicação e Cultura**; v. 17, n. 33, 2018.

VIŠNOVSKÝ, J; UNGEROVÁ, M; KUBÍKOVÁ, K. **Robo-Journalism and Its Implementation in Editorial Practice**. NORDSCI, 2019.