



CEST

Centro de Estudos Sociedade e Tecnologia



Universidade de São Paulo

Boletim- Volume 8, Número 07, Agosto/2023

ChatGPT e outros LLMs: Estamos na Era da Quase-Verdade?

Edison Spina, Mario Magalhães & Marcel Simonette

Introdução

Você acompanharia um texto que boa parte soa como uma repetição de tautologias presentes em várias fontes ou parte do senso comum? Construiria sobre referências bibliográficas sem garantir se são reais, se os autores e trabalhos realmente existem ou foram publicados? Confiaria em uma planilha que, ao calcular gerasse aleatoriamente alguns valores falsos, mas que parecessem corretos?

Cada vez mais se encontram informações, em diversas formas e meios: textos, imagens, vídeos, ou dados que parecem reais e corretos, mas, muito além das *fake news* forjadas por indivíduos, são geradas por sistemas, algoritmos, tecnologias de inteligência artificial criadas para parecer... “inteligentes”.

Facilmente se pode acreditar nelas, são “quase-verdades”.

Parecem verdadeiras, e parte delas são, mas não todas elas. Aparentemente são geradas com inteligência e cultura. Convencem pela sua forma e apresentação.

A Era da quase-verdade

Pode-se estabelecer que se entrou na era da quase-verdade. Um período em que, a informação, independente da forma ou meio, é abundante, está disponível e o acesso é simplificado. A informação é aparentemente verdadeira, mas muitas vezes conflitante. As fontes, mesmo que declaradas também podem ser forjadas na forma e estilo de escrita e apresentação.

É necessário esforço significativo para validar o que é verdadeiro e o que é forjado, dependendo de conhecimento prévio ou direto, de acesso a informações já validadas e de interpretações consensadas entre os validadores. Porém, o excesso de fontes, a intrincada rede de referências potenciais, a falta de legitimidade tanto dos geradores, quanto dos processos de validação impõem baixa confiabilidade na informação e ajudam a manter desinteressados, ou mal-informados, seus consumidores, que encontram um conforto relativo ao ficarem alheios a esses problemas.

A sociedade encontra-se diante deste desafio em larga escala.

Como se chegou a esse ponto?

Nos primórdios da sociedade, a natureza era para o homem a representação da verdade, e os deuses, inquestionáveis, eram forças da natureza que tudo determinavam harmoniosamente. Esta visão encontra-

se bem representada na filosofia grega, em particular, aristotélica.

A evolução da sociedade levou a religiões organizadas, dogmáticas e baseadas na fé. A verdade assim, claramente estabelecida, determinava corolários de compreensão fácil e aplicação absoluta.

Com o Iluminismo, a verdade era a solução encontrada por processos bem controlados de explicações dos eventos. Acreditava-se que, se algo não fosse entendido e explicável, seria apenas necessário mais conhecimento para que a verdade ficasse estabelecida. Elementos, no entanto, que continuavam a gerar dúvidas precisavam ser verificados. Postulados e teorias eram formulados e confirmados pela comunidade com replicabilidade e autenticação de resultados.

Da certeza da verdade dos conceitos iluministas, cientistas como Newton realizaram todo o seu

A incerteza da informação passa a ser a solução dos problemas não resolvidos pelo Iluminismo



desenvolvimento de teorias. Alguns problemas, porém, se tornaram demasiadamente complicados, e as teorias começaram a ser questionadas assim como seus postulados.

O entendimento e a incorporação da incerteza formam os avanços marcantes realizados por diversos pensadores no início do século XX, entre eles Gödel e Heisenberg. A incerteza da informação passa a ser a solução dos problemas não resolvidos pelo Iluminismo. A admissão da dúvida foi a grande vitória conceitual dessa fase. Novas lógicas foram desenvolvidas para incluir o fato, a negação do fato e a incerteza sobre o fato como parte da solução dos problemas e dos modelos.

Simultaneamente, ainda que limitada mundialmente pelas formas de comunicação, a disponibilidade de informação alcança dimensões globais. A imprensa, os jornais, posteriormente o rádio e a televisão alcançam a maioria da população. E, tacitamente, o que informam é aceito como verdadeiro.

Nas últimas décadas do século XX depara-se com um novo desafio, uma verdade de consumo, que aparenta não precisar ser validada ou é difícil de ser verificada, mas que acomoda as dúvidas e que são assim consumidas com facilidade. Nesse conceito, a comunicação de massa passa a fazer parte importante do processo de criação de narrativas acomodando fatos de difícil explicação, omitidos ou desconhecidos para atender determinados interesses não expressos.

A geração de narrativas é a base da pós-verdade. A narrativa montada a posteriori para se consumir a informação como uma versão verdadeira ao invés do aprofundamento em busca dos fatos geradores. A pós-verdade gera versões desinfetadas de conceitos não convenientes ou desejados para alguma das partes envolvidas diretamente ou não. A supressão da verdade que estaria sendo verificada através de uma estória aparentemente verdadeira, mas que contém os vieses convenientes para difusão e profusão.

A quase-verdade vem a ser uma evolução inesperada na comunicação da informação pela sua geração de

discursos ou narrativas com auxílio de recursos de tecnologia, principalmente o que se denominou Inteligência Artificial. As narrativas são construídas com um aparente suporte de referências, conceitos históricos, lógica de aplicação e suporte bem estruturadas, em texto bastante palatável e indiscutível na forma, criando uma verdade aparente.

A construção de quase-verdades

A geração de textos e narrativas quase-verdadeiras é resultado de um processo lógico, mas não de uma inteligência com algum conceito semântico. É uma composição estatística a partir da análise de sucessões de palavras similares em bases massivas de textos transformados em representações matemáticas independentes de sua semântica. São textos ou narrativas que formam uma quase verdade leviana, destituída de interpretação semântica, que ganha sentido ou porque reflete as tautologias de maior frequência ou porque reúne palavras que formam

conceitos que probabilisticamente apareceriam em textos semelhantes. Quando questionados, não se encontra suporte real ou formal. A forma passa a ser importante para impor uma aparente estrutura conceitual.

Adicionalmente as soluções hoje disponíveis incluem uma camada final de filtro ou crítica visando evitar que os textos gerados incluam assuntos polêmicos ou ilícitos. Estes filtros refletem a polarização de quem os controla.

Os textos gerados automaticamente por estes algoritmos, com palavras ou sequências de palavras advindas de toda a história de textos disponíveis, e até criando aparentes referenciais teóricos, são superficiais e levam a grande quantidade de consumidores ávidos por explicações fáceis e palatáveis, com vieses convenientes para a narrativa pretendida que apostam no espaço de conforto do consumidor.

A quase-verdade pode levar um conjunto de cidadãos a acreditar em fatos devidamente polarizados sobre qualquer assunto e o poder, ou a política, vai ocupar-se de prevalecer, segundo as regras dos criadores de regras de “pensamento” da IA. O aprendizado de máquina

O ChatGPT não depende de conhecimento prévio, mas consegue entender e manter o contexto em um diálogo



supera facilmente o aprendizado criativo e investigativo em volume, velocidade e disponibilidade. O conforto do cidadão vai ser o grande apoiador do crescimento da quase-verdade.

Convivendo com a realidade da quase-verdade

Um exemplo de um destes sistemas é o chatGPT da OpenAI. É um sistema utilizando soluções de IA baseado em modelos conhecidos como LLM (*Large Language Models*) com uma camada de filtro associada capaz de dialogar com os usuários de maneira muito natural.

O ChatGPT é um produto visando o público em geral, não depende de conhecimento prévio, mas consegue entender e manter o contexto em um diálogo. Pode, com formas muito simples, responder às questões colocadas e trazer respostas claras em linguagem natural. Constrói suas respostas a partir do conteúdo da Internet usado para treinar o sistema.

Não há aprendizado real no sistema. O aprendizado tem a função de criar conhecimento e experiências que supram a capacidade de solução dos problemas a serem resolvidos nos seus diversos contextos semânticos: técnicos, humanos, de trabalho ou de lazer. O sistema desconhece a carga semântica e seus contextos. É feita uma acumulação de estatísticas relacionadas na construção de mapas que representam presenças, semelhanças e proximidades ou temperatura (uma medida de impacto ou frequência da palavra nos conjuntos de mapas) de palavras. Não há desenvolvimento de competência na solução de problemas, pois não há acúmulo de experiências. Há um treinamento baseado no acúmulo de fragmentos de textos similares ou relacionados.

O sistema conta com o conhecimento global, mas é organizado de acordo com diretrizes traçadas pelos modelos estatísticos e seus parâmetros. Como ocorria no passado com a Wikipedia, onde os verbetes poderiam ter seus textos escritos por qualquer um e contendo qualquer polarização sem serem, necessariamente, corrigidos ou monitorados. As informações incorporadas são adquiridas pelos modelos com baixo nível de coordenação ou curadoria. As diretrizes que são utilizadas no mecanismo de IA

não são rastreáveis ou, tão pouco, controláveis. Quem cria essas diretrizes são elementos desconhecidos e podem polarizar respostas, segundo suas próprias crenças ou interesses. Há bilhões de diretrizes catalogadas no sistema e não há forma conhecida ou possível para garantir que sua destreza nas respostas não esteja dependendo apenas de conceitos com maior ocorrência, sem explicações e fundamentos ou, o que seria mais perigoso, de conceitos devidamente criados para controlar narrativas. Não há referências confiáveis para quaisquer informações disponibilizadas. O controle do processo não existe, ou, é tão distribuído que não pode ser imputado a ninguém.

É sim possível utilizar a IA para recuperar informações até verdadeiras, mas saber perguntar e confiar na “opinião” do resultado obtido será, cada vez mais, uma arte.

Não basta perguntar, é necessário formular para os sistemas de IA o contexto e a pergunta correta. Não basta receber a resposta é preciso ser capaz de duvidar e questionar o recebido.

A sociedade em geral e os indivíduos em particular estão preparados?



Edison Spina é coordenador acadêmico do CEST.



Mario Magalhães é pesquisador do CEST



Marcel Simonette é pesquisador do CEST

Este artigo resulta do trabalho de apuração e análise dos autores, não refletindo obrigatoriamente a opinião do CEST.