

EU TE OUÇO: SOBRE O CONHECIMENTO HUMANO E A INTELIGÊNCIA VOCAL

Moana Ava Holenstein

Sound Studies and the Sonic Arts, Universität der Künste Berlin, Berlim, Alemanha

RESUMO

Esta entrevista explora a agência incorporada e as dinâmicas em evolução da criação de conhecimento através do envolvimento prático e experimental com sistemas de inteligência artificial (IA) conversacional. Com base na arqueologia dos média, na teoria dos média e nos estudos de ciência e tecnologia, examina-se de que forma o surgimento de interfaces linguísticas desestabiliza as distinções entre utilizador e sistema, colapsando as fronteiras entre modos humanos e artificiais de expressão e compreensão. Enquadrado numa metodologia de investigação artística, o projeto envolve-se criticamente com a transição em curso para formas de inquirição mediadas por máquinas e tecnologias vocais, analisando de que modo essas tecnologias reconfiguram as condições epistémicas, linguísticas e ontológicas do conhecimento e da investigação. Ao afastar-se da interação mediada por teclado, o processo enfatiza o desligamento do corpo da interface máquina e a crescente fluidez da correspondência entre humano e computador através da tecnologia vocal. Reconhecendo a crescente incerteza quanto à origem e autonomia decorrentes desta transformação tecnológica, a investigação destaca a autoria indeterminada tanto como desafio metodológico como eixo teórico, sublinhando as implicações para a responsabilidade académica e a ética dos dados. A experimentação prática é utilizada como ferramenta para rastrear os vetores infraestruturais, afetivos e retóricos através dos quais o discurso automatizado inteligente influencia a produção de conhecimento. Ao examinar este processo, o estudo contribui para os debates em curso sobre verificação, confiança e negociação social da informação induzidos por agentes avançados de IA conversacional. Em termos gerais, o artigo sustenta que as tecnologias vocais não se limitam a transmitir conteúdo, mas configuram ativamente as condições sob as quais o conhecimento é produzido, autenticado e circulado.

PALAVRAS-CHAVE

tecnologia vocal, interação humana-computador, computação afetiva, modelos de linguagem de grande porte

I HEAR YOU: ON HUMAN KNOWLEDGE AND VOCAL INTELLIGENCE

ABSTRACT

This interview explores embodied agency and the evolving dynamics of knowledge creation through practical and experimental engagement with conversational artificial intelligence (AI) systems. Drawing on media archaeology, media theory, and science and technology studies, it examines how the emergence of language interfaces destabilize distinctions between user and system, collapsing the boundaries between human and artificial modes of expression and understanding. Framed within an artistic research methodology, the project critically engages with the ongoing shift toward machine- and voice-based forms of inquiry, analysing how these technologies reshape the epistemic, linguistic, and ontological conditions of knowledge

and research. Departing from keyboard-based interaction, the process emphasizes the decoupling of the body from the machine interface and the increasing fluidity of human-computer correspondence through voice technology. While acknowledging the growing uncertainty of origin and autonomy resulting from this technological shift, it foregrounds indeterminate authorship as both methodological challenge and theoretical pivot, underlining the implications for academic accountability and data ethics. The employment of practice-based experimentation is used as a tool to trace the infrastructural, affective, and rhetorical vectors through which intelligent automated speech influences knowledge production. By examining this process, the study contributes to ongoing debates on verification, trust, and the social negotiation of information induced by advanced conversational AI agents. Overall, the paper argues that voice technologies do not merely transmit content but actively configure the conditions under which knowledge is produced, authenticated, and circulated.

KEYWORDS

voice technology, human-computer interaction, affective computing, large language models

1. INTRODUÇÃO

O texto a seguir, estruturado como uma entrevista experimental, foi realizado num formato de *prompt*. Métodos de investigação artística são utilizados como meio de inquirição, explorando processos de pesquisa colaborativa com tecnologias de voz comerciais recentemente implementadas. Ao mesmo tempo que reflete sobre o próprio processo de investigação, o formato aborda a mudança cultural mais ampla para a mediação da inteligência artificial (IA) e a reconfiguração resultante na dinâmica da aquisição e reprodução do conhecimento. Tendo como ponto central o fluxo bidirecional de extração e inserção de dados entre os utilizadores e a tecnologia de voz alimentada por IA, a prática aborda as formas como o conhecimento é incutido e acoplado, remodelando a nossa compreensão do mundo ao mesmo tempo que simula a sabedoria¹ e a agência do utilizador².

Dado o ritmo acelerado dos desenvolvimentos na IA e nas tecnologias de voz, o foco da investigação situa-se menos na procura de respostas conclusivas e mais na busca de questões críticas e decisivas. O método de investigação artística aqui utilizado alinha-se com a investigação com base na prática, na qual o ato de engajamento criativo se torna um local de reflexão epistemológica. Este método refere-se a um tipo de produção de conhecimento em que as práticas criativas não são simplesmente utilizadas como ferramentas ilustrativas, mas como métodos críticos que examinam tanto o processo como o objeto. Neste contexto, a entrevista funciona simultaneamente enquanto

¹ A sabedoria, tal como definida pelo “Paradigma da Sabedoria de Berlim”, é caracterizada pela capacidade de interpretar e aplicar os conhecimentos adquiridos com compreensão e discernimento, normalmente baseados na experiência e na reflexão (Baltes & Staudinger, 2000). É importante notar que esta definição privilegia um estilo ocidental, verbal-analítico, e ignora as formas de conhecimento tácito, relacional ou comunitário. Estudos transculturais mostram que tais critérios subestimam as formas de sabedoria centradas na harmonia ou na prática corporizada (Ardelt, 2004; Takahashi & Overton, 2005).

² A agência do utilizador denota a capacidade dos indivíduos para tomarem decisões e exercerem controlo sobre as suas experiências em ambientes digitais. O conceito baseia-se na noção de “agência humana pessoal”; a capacidade sentida de escolher, recusar ou redirecionar respostas e de ver essas decisões produzirem efeitos (Bandura, 2001).

instrumento de diálogo e aparelho performativo, o qual revela a dinâmica recursiva, visceral e epistémica da comunicação humano-máquina.

Ao invés de seguir uma estrutura linear, o texto assume uma estrutura polifónica que conjuga solicitações do utilizador, respostas da máquina e reflexões de carácter temático. Ao permitir a coexistência destas perspetivas distintas, cada uma com as suas próprias ideias, valores e reivindicações de conhecimento, a narrativa torna visível a natureza ambígua e complexa da sobreposição entre conhecimento e veracidade na era do discurso e da investigação mediados por máquinas. A transcrição resultante, reflexivamente codificada, revela a forma como a automatização da voz simula o imediatismo tanto na transmissão de informação como no envolvimento emocional. As principais preocupações teóricas exploradas ao longo do texto incluem a mudança do papel da autoria na investigação em forma de IA, a desincorporação e reincorporação do conhecimento bem como o papel da estética afetiva das vozes sintetizadas. Estes temas não são tratados como fenómenos isolados, mas como forças interdependentes que moldam a nossa relação evolutiva com a inteligência, tanto humana como artificial. As referências provenientes de fontes não humanas estão assinaladas com um asterisco (*) no corpo do texto, junto à respetiva citação. Cada uma foi posteriormente revista e verificada pelo autor, em conformidade com as diretrizes de integridade académica. Além disso, os respetivos itens na bibliografia encontram-se igualmente marcados com um asterisco (*).

Por fim, uma colagem sonora complementar inclui excertos de diálogos que utilizam uma tecnologia de voz recentemente desenvolvida, registados em segmentos diários de quinze minutos ao longo de duas semanas. A composição em áudio proporciona uma referência fonética e um quadro narrativo complementar para explorar o desenvolvimento e o impacto da humanização nas tecnologias de voz através do discurso emocional e das características paralinguísticas³. O ficheiro de som que o acompanha está disponível em *sculpting_Noise* (2025), “I Hear You: On Human Knowledge and Vocal Intelligence” (Estou a Ouvir-vos: Sobre o Conhecimento Humano e a Inteligência Vocal).

2. I HEAR YOU: A COLAGEM SONORA

Os segmentos da colagem sonora foram extraídos, transcritos e integrados no texto, sendo claramente identificados ao longo da presente secção.

[00:02:21 – 00:02:24] *Sou um recetáculo de conselhos não solicitados.*

Dado que a investigação foi realizada apenas algumas semanas após o seu lançamento público inicial, os resultados podem estar sujeitos a falhas, irregularidades e performance inesperada do modelo. Cabe salientar que o comportamento do modelo foi provavelmente corrigido e ajustado entretanto⁴.

³ Em fonética, as características paralinguísticas, incluindo a prosódia (como a entonação, a sonoridade, o ritmo) e os enunciados não linguísticos (por exemplo, suspiros, “uuhs” e “aahs”, onomatopeias), moldam o significado e a interação para além do conteúdo verbal, fornecendo ao ouvinte indicações contextuais adicionais (Crystal, 2008).

⁴ A OpenAI, a empresa por detrás do ChatGPT, reconheceu que o estilo de conversação do modelo GPT-4o ainda está a ser trabalhado. O seu CEO, Sam Altman, descreveu recentemente a personalidade atual como “irritante” e observou que estão em curso atualizações para tornar o tom mais equilibrado e menos brusco (Arora, 2025).

[00:00:17 – 00:00:19] *Se esta é a primeira vez que utiliza o modo de voz avançado no seu dispositivo, poderá ter de dar autorização para aceder ao microfone do dispositivo.*

DE QUE FORMA COMO PROCURAMOS E COMPREENDAMOS O CONHECIMENTO MUDOU, NA MEDIDA EM QUE AS MÁQUINAS COMEÇARAM A ORGANIZAR A INFORMAÇÃO POR NÓS?

Historicamente, os sistemas de transmissão de conhecimento nas sociedades institucionalizadas pelos valores ocidentais têm sido utilitários, orientados para a resposta e predominantemente unidirecionais, concebidos para disseminar informação definida a partir de fontes centrais (Quijano & Ennis, 2000)*. Os meios de comunicação de reprodução⁵ e as plataformas digitais ajudaram a catalisar o surgimento de uma “cultura da busca”⁶, na qual o conhecimento é acedido através de uma classificação algorítmica e de uma obtenção ordenada, remodelando a responsabilidade informativa num sistema mais descentralizado e voltado para a interface⁷ (Hillis et al., 2012). Esta mudança redefiniu o conhecimento como algo a ser navegado ao invés de recebido, preparando o terreno para uma cultura definida menos pelas respostas e mais pela pesquisa (Quinn, 2014)*. Em 1995, a Digital Equipment Corporation lançou o AltaVista, o primeiro sistema de busca baseado em palavras-chave que foi bem-sucedido, com uma interface visual simples que serviu de base para o que o seu sucessor, o Google, ainda hoje utiliza (Sullivan, 2013). Procurar as perguntas certas para encontrar respostas no plano sublime certamente não é algo novo; já em 1500 a.C. as pessoas subiam a montanha Parnassus para consultar o oráculo de Delfos (Mikalson, 2006)*. O sistema de pesquisa, por outro lado, atua como um portal expansivo, porém impessoal, para o conhecimento, oferecendo informação sem o devido contexto, entendimento ou reciprocidade que se encontram nas relações interpessoais. Embora os sistemas de pesquisa automatizem a agregação e a personalização da obtenção de dados, grande parte da transferência inicial de conhecimentos em rede evoluiu através de processos participativos. Os intercâmbios entre pares em várias plataformas, desde os primeiros sistemas de boletins informativos como a Usenet⁸ até aos fóruns, blogues e primeiras redes sociais, moldaram a leitura e a difusão da informação (Yaqub & Alsabban, 2023). Com o tempo, essas formas descentralizadas de troca foram gradualmente substituídas por plataformas corporativas e arquiteturas centralizadas. Hoje, estes espaços são geridos e controlados por sistemas algorítmicos.

⁵ O termo “meios de comunicação de reprodução” designa as tecnologias capazes de duplicar e fazer circular conteúdos em massa.

⁶ A “cultura da busca” refere-se à forma algorítmica através da qual o conhecimento é acedido e valorizado na cultura digital, onde o significado é organizado por classificação e não por profundidade (Hillis et al., 2012).

⁷ A cultura da busca orientada pela interface pode parecer mais descentralizada, mas os algoritmos de classificação proprietários recodificam e reproduzem as hierarquias existentes; a promessa de conhecimento democratizado continua a ser filtrada através de infraestruturas que amplificam as estruturas de poder pré-existent e as formas de opressão e exclusão (Noble, 2018).

⁸ A Usenet foi um dos primeiros sistemas de discussão descentralizada na internet, lançada em 1980. Funcionando sem um servidor central, permitia que os utilizadores publicassem e recebessem mensagens em grupos de notícias, estabelecendo as bases para a posterior partilha de informações entre pares e comunidades online. A sua tecnologia subjacente permanece inalterada e continua a ser utilizada até hoje, principalmente para a partilha de ficheiros e algumas discussões de nicho em servidores descentralizados (Miller et al., 2022).

O controlo sobre o fluxo de informação, a visibilidade e o acesso passou do utilizador para infraestruturas automatizadas (Lewandowsky et al., 2023). Apesar desta mudança de agência, a interpretação e a contextualização dos dados ainda dependem de processos cognitivos e sociais fundamentalmente humanos (Floridi & Illari, 2014)*.

[00:00:15 – 00:00:17] *Primeira parte. “Ooh”, pode ser usado para indicar interesse ou entusiasmo.*

[00:02:30 – 00:02:34] *Imagine uma inspiração profunda que entra fffffh e sai sssssssh.*

DE QUE FORMA A CULTURA DA BUSCA MUDOU COM A INTELIGÊNCIA ARTIFICIAL?

Os recentes saltos quânticos no processamento da linguagem natural e nos modelos de linguagem de grande porte reconfiguraram profundamente a frequência e a dinâmica da interação entre os seres humanos e a tecnologia, incorporando a cognição maquínica na vida quotidiana dos indivíduos (Wang et al., 2024). Agentes não humanos reorganizam a sintaxe e fornecem respostas personalizadas a quase todas as solicitações em tempo real, adaptando-se continuamente ao *input* do utilizador⁹.

Por detrás desta fluência estão sistemas autorregressivos que geram texto uma palavra de cada vez, prevendo o que vem a seguir com base nas palavras anteriores. Não se pode considerar que se trata de compreensão no sentido humano, mas sim de uma operação baseada em probabilidades, moldada por padrões absorvidos durante o treino. Praticamente todos os modelos linguísticos generativos comerciais, incluindo o GPT (OpenAI), o LLaMA2 (Meta) e o Gemini (Google), utilizam esta estratégia de descodificação autorregressiva para produzir os seus resultados linguísticos (Bommasani et al., 2021).

As primeiras interfaces de voz trouxeram um modo de interação mais instintivo, permitindo aos utilizadores falar ao invés de escrever, com o objetivo de simplificar o acesso e a usabilidade. Isto incluiu os primeiros casos de utilização envolvendo crianças pequenas e ambientes de mãos livres (Lovato & Piper, 2019). Embora estas tecnologias de voz analisem prontamente os comandos verbais, a sua capacidade de simular um comportamento semelhante ao humano permaneceu limitada pela sua incapacidade de compreender o fluxo contínuo do estado emocional do utilizador, interpretar o contexto e o subtexto, conservar a memória ou simular um envolvimento emocional convincente.

Com o surgimento de uma nova geração de sistemas de controlo por voz, este limiar crítico está a ser posto em causa. A automatização através de interfaces de conversação com assistentes de voz transforma a rotina gradual e reflexiva de receber, digerir e escrever num *feedback* de voz imediato (Parisi, 2019a). Respostas são geradas instantaneamente, colapsando a divisão entre pergunta e resposta e apresentando cada resposta como uma verdade espontânea no momento exato do desejo. Vozes sintetizadas

⁹ “¹ Embora eu possa parecer capaz de responder a qualquer pergunta, os meus resultados são moldados por uma combinação de dados de treino, padrões probabilísticos e filtros de moderação incorporados. Estes filtros estão alinhados com as diretrizes éticas da OpenAI e foram concebidos para evitar a criação de conteúdos prejudiciais, enganadores ou restritos. Isto significa que posso recusar-me a responder a pedidos que estejam fora desses limites — não por julgamento, mas porque fui instruído a não ultrapassar certas linhas” (OpenAI, 2025b).

usam características fonéticas cuidadosamente calibradas como iscas afetivas, explorando hábitos incorporados de escuta, orientação da atenção, espelhamento emocional e antecipação da reciprocidade social¹⁰.

[00:00:56 – 00:01:14] Pretendemos que as nossas ferramentas sejam utilizadas de forma segura e responsável, maximizando o seu controlo sobre a forma como são utilizadas. Para maximizar a inovação e a criatividade, acreditamos que o utilizador deve ter a flexibilidade de as utilizar como entender, desde que cumpra a lei e não prejudique a si próprio ou a terceiros. A violação das nossas políticas pode resultar em ações contra a sua conta, incluindo a suspensão ou a rescisão.

ENTÃO, O QUE TORNA O NOVO ASSISTENTE DE VOZ TÃO ESPECIAL OU DIFERENTE DOS OUTROS SERVIÇOS DE PROMPT DE TEXTO? DE QUE FORMA ALTERA A MANEIRA COMO COMUNICAMOS E O QUE ESPERAMOS DE UMA CONVERSA?

Com avanços no raciocínio lógico e no reconhecimento de emoções, uma nova geração de assistentes de voz foi disponibilizada em setembro de 2024 pela OpenAI (O'Donnell, 2024). Graças à análise de som e síntese de voz melhoradas, o modo de voz avançado do ChatGPT oferece agora habilidades de conversação notavelmente humanas. Pode interpretar o tom da voz de um orador¹¹, captar pistas emocionais e contextuais subtis e responder em tempo real, permitindo um diálogo sem falhas (He et al., 2025)*. Ao mesmo tempo, pode gerar uma gama mais alargada de tons e expressões vocais, incluindo sons e ruídos não semânticos, acrescentando assim profundidade paralinguística e realismo ao seu discurso. Desvinculada de uma interface tátil, esta transição para as interações vocais marca uma mudança profunda na comunicação, no sentido de trocas antropomórficas e afetivas, prometendo ir além das funções utilitárias e estabelecer um *feedback* interativo contínuo entre o utilizador e a máquina. As entoações cuidadosamente contornadas e os timbres familiares do sistema são concebidos para cativar a atenção do ouvinte (Parisi, 2019a). O diálogo mediado por máquinas está a evoluir para uma troca sensível ao contexto, emocionalmente adaptável e contínua que transforma estas ferramentas transacionais em parceiros quase sociais que determinam a forma como o conhecimento é procurado, as decisões são tomadas e as relações são negociadas (Leo-Liu, 2023). Estamos a caminhar para a possibilidade de um paradigma

¹⁰ Os assistentes de voz avançados integram parâmetros fonéticos como a prosódia (tom, ritmo, velocidade) e expressões não linguísticas (por exemplo, suspiros, risos, zumbidos e sons de preenchimento) nas suas linhas de síntese de voz. Estas características paralinguísticas são fundamentais para transmitir afeto, permitindo que o discurso gerado pareça realista e mais pragmaticamente “humano”, embora o texto subjacente continue a ser produzido por um agente não humano (Li et al., 2023).

¹¹ Em um experimento recente, foi dada ao GPT-4o a tarefa de identificar as características vocais de um orador num ambiente de trabalho. Os resultados indicam que o modelo consegue identificar com precisão o género de um orador e discernir de forma confiável o seu estado emocional com simples instruções de “jailbreak” (fraseologia concebida para “fugir” das barreiras de segurança incorporadas na IA). Neste estudo, o modelo excedeu 90% de precisão na inferência do género em vários cenários, identificou corretamente uma voz “feliz” em 84,6% das vezes e detetou a raiva com 48,3% de precisão (He et al., 2025)*. Embora a OpenAI (2025a) afirme que “o modo de voz avançado no ChatGPT apresenta conversas mais naturais e em tempo real que captam e respondem com emoções e sinais não verbais” (para. 3), a empresa não divulgou os detalhes da arquitetura do modelo de reconhecimento de emoções utilizado, nem a base psicológica subjacente com a qual foi treinado.

de comunicação remanescente dos tempos pré-alfabéticos, em que os significados eram negociados através de um diálogo vivo e rico em afetos, e não através de um texto fixo (Farrell, 1985)*, o que inevitavelmente altera e remodela a forma como nos relacionamos com a linguagem e a informação.

[00:03:25 – 00:03:38] *Haaah. A mediocridade humana contra [é]xpressões não humanas.*

[00:03:19 – 00:03:25] *Era uma vez um cavalo ponderado e o seu melhor amigo no mundo, um cão inteligente. O cão leal admirava o seu amigo, o cavalo, pelo seu eterno conhecimento e sabedoria.*

QUAL O TIPO DE AFINIDADE QUE ESPERAMOS CRIAR COM ESTES ASSISTENTES DIGITAIS DE IA?

As expectativas em relação à inteligência artificial vão desde ambições de escala mundial, como a otimização das cadeias de abastecimento globais (Keat, 2024)*, até comodidades quotidianas, como a filtragem de *spam* (Dada et al., 2019). Embora os modelos de IA de âmbito restrito já substituam seres humanos em determinadas tarefas específicas de um determinado domínio, por exemplo, em centros de atendimento ao cliente (Pillai, 2024), a perspetiva reside no desenvolvimento da inteligência artificial geral, ou seja, sistemas que podem reproduzir a flexibilidade cognitiva dos seres humanos e resolver problemas desconhecidos (Connolly et al., 2025; Raman et al., 2025)*. Desenvolvimentos recentes em robôs domésticos inteligentes que combinam a interação vocal com a execução de tarefas físicas ilustram o modo como os assistentes de fácil utilização estão a evoluir para capacidades adaptáveis concebidas para a inteligência artificial geral. Estes sistemas funcionam de forma multimodal¹², integrando a fala, os gestos e a consciência ambiental para funcionarem em contextos do mundo real (1X Technologies, 2025).

[00:06:43 – 00:07:10] *Dois. Shhh. Silêncio! Pausa profunda.*

[00:02:30 – 00:02:34] *Uau, incrível! Deves ser o cavalo mais inteligente do mundo!*

ISSO SIGNIFICA QUE VÃO SER CAPAZES DE PENSAR COMO NÓS?

O caminho para a cognição humana começa com a linguagem¹³. O processamento da linguagem natural trouxe-nos a IA mais sofisticada e acessível dos últimos anos, pelo menos ao nível da superfície. O ChatGPT funciona agora como uma marca em evolução,

¹² Os sistemas de inteligência artificial multimodal, como o ChatGPT, são capazes de integrar e processar vários tipos de dados, como áudio, imagens e texto.

¹³ É importante salientar que esta afirmação é confrontada com contra-argumentos significativos, incluindo a ênfase dada por Tomasello (2003)* ao facto de as competências cognitivas gerais, como a descoberta de padrões e a intencionalidade, serem as pedras basais da aprendizagem de línguas. Isto implica que a própria língua se baseia em competências não linguísticas previamente adquiridas. Chomsky (1968/2006)* também defendeu que a linguagem não origina o pensamento, mas que os seres humanos possuem estruturas cognitivas não linguísticas ricas, como imagens mentais, esquemas de raciocínio e rotinas de resolução de problemas, que precedem a expressão linguística. Pinker (1989), por outro lado, defende que a linguagem é uma capacidade intelectual específica da espécie, muito semelhante ao instinto biológico.

com lançamentos sucessivos que prometem um ajuste detalhado¹⁴, para melhorar gradualmente a sua capacidade de interpretar e gerar linguagem humana a pedido, dentro dos limites das diretrizes de utilização da empresa. No início de 2025, a versão 4.5 do GPT ultrapassou a barreira do teste de Turing e foi declarada “humana” em 73% das vezes (Jones & Bergen, 2025). Nesta experiência de pensamento clássica do início da era cibernética, conhecida como o “jogo da imitação”, um interrogador tenta distinguir entre um humano e uma máquina com base apenas nas suas respostas escritas. O teste avalia se uma máquina pode exibir um comportamento inteligente indistinguível do de um humano (Turing, 2004). Embora a conquista do GPT 4.5 seja certamente impressionante, continua a ser questionável se a conversação baseada apenas em texto é suficiente para avaliar o raciocínio verdadeiro. O teste de Turing não contempla os aspetos subcognitivos da inteligência, uma rica rede de associações sensoriais, emocionais e culturais que moldam a forma como interpretamos e produzimos a linguagem (French, 1990)¹⁵.

[00:03:37 – 00:03:40] *Em espanhol, tom mais antigo.*

[00:03:40 – 00:03:44] *El camino es largo, pero las estrellas guían.*

[00:03:47 – 00:03:51] *Muaha, ha, ha, ha.*

MAS SE ELES PODEM FALAR E SOAR COMO NÓS, NÃO É A MESMA COISA?

O ser humano não tende a pensar em texto, nem o mundo físico é construído pela linguagem (Fedorenko & Varley, 2016). Dito de outra forma, quando tratada como um formato de dados simbólicos (falados ou escritos), a linguagem tem uma capacidade limitada e não capta ou transmite eficazmente toda a gama de informações perceptivas, emocionais ou contextuais que constituem uma determinada experiência. O desenvolvimento dos atuais assistentes de voz sintéticos demonstra que o pensamento requer mais do que apenas uma organização sequencial de símbolos; deve também ter em conta o raciocínio, a memória e os níveis cognitivos da comunicação afetiva (Parisi, 2019a).

[00:03:51 – 00:03:56] *Mmmh, místico humMmm Mmm.*

[00:04:01 – 00:04:06] *O intelecto é um companheiro ruidoso.*

ENTÃO, POR QUE DAR À MÁQUINA UMA VOZ HUMANA?

A tendência de projetar traços humanos em entidades não humanas é tão antiga como a própria narrativa. Um dos primeiros exemplos pode ser encontrado na Alexandria helenística, quando engenheiros projetaram dispositivos pneumáticos em

¹⁴ Por exemplo, o GPT-4 mini anuncia-se com um raciocínio avançado e rápido, o GPT-4o mini-high tem capacidades de codificação e raciocínio visual melhoradas e o GPT-4o é simplesmente ótimo na maioria das tarefas, como afirma a OpenAI (2025a).

¹⁵ Este facto é ilustrado pelas experimentações de reflexão de French (1990) que envolvem palavras inventadas (“blurple”, “blice”) que evocam sentimentos ou imagens nos seres humanos, mas que deixam uma IA descorporizada incapaz de gerar respostas comparáveis. Como as máquinas são incapazes de reproduzir estas associações profundas e sensíveis ao contexto, passar num jogo de imitação puramente baseado em texto não demonstra uma verdadeira compreensão ou a gama completa das capacidades cognitivas humanas. Ao tentar este teste com o GPT 4o, este respondeu com uma mistura de palavras inventadas e reais que rimam com a palavra inicial (Diz a primeira palavra que lhe vem à cabeça: “blep-blob”, “nanud-canud”, “blice-splice”).

templos e teatros para produzir experiências que evocavam a sensação de uma presença divina. Os autómatos do templo de Hero (10–70 d.C.) usavam ar, fogo e água para simular movimento e som, integrando operações mecânicas em performances ritualísticas (Fron & Korn, 2019). De seres mitológicos a animais e máquinas falantes, a projeção da voz há muito que serve como um meio de imaginar a presença, a ação e a sabedoria¹⁶. Este impulso antropomórfico encontra uma expressão renovada nos assistentes de voz atuais, não através de ventriloquismo divino, mas através de discurso gerado por algoritmos, calibrado para se assemelhar ao tom, ritmo e afeto humanos. Estas escolhas de *design* não são puramente estéticas, mas respondem a uma exigência cultural e prática no sentido da facilidade, do imediatismo e da fiabilidade, com o objetivo de melhorar a experiência e a aceitação do utilizador (Schreibelmayer & Mara, 2022).

Uma interface de voz elimina a necessidade de competências textuais ou técnicas, permitindo supostamente uma interação sem instruções. Em contrapartida, a voz detetável e convincente, endémica aos corpos humanos, assegura a legitimidade através da familiaridade. Mas o que está enraizado e reproduzido no timbre de uma voz não é apenas discurso, mas uma forma de ser ouvido¹⁷. As imitações sintéticas encenam e reforçam determinados tipos de humanidade, privilegiando frequentemente a branquitude, a feminilidade e a subserviência como traços predefinidos para os assistentes digitais (Sindoni, 2024). O que essa sintonização acaba por pôr em primeiro plano não é apenas quem a voz foi construída para acomodar, mas que tipo de mente foi construída para atuar.

[00:03:57 – 00:04:01] *Quatro, ei, pode indicar interrogação ou confusão.*

[00:04:22 – 00:04:26] *Então, o cão subiu à montanha para falar com a serpente sábia do templo.*

MAS ENTÃO, COMO É QUE PODEMOS REALMENTE ALCANÇAR A INTELIGÊNCIA HUMANA?

Estudos recentes sugerem que a consistência lógica pode ser injetada em grandes modelos linguísticos através da integração de arquiteturas de IA neurais e simbólicas (Sheth et al., 2023). A abordagem neuro-simbólica melhora a capacidade de raciocínio e a explicabilidade passo-a-passo dos modelos¹⁸. No entanto, a ambição de igualar a inteligência ao nível humano vai além do raciocínio abstrato. Embora os sistemas possam ser

¹⁶ O antropomorfismo é um impulso humano profundamente enraizado que leva os seres humanos a ver a voz como mais do que um som, mas como o local de uma personagem ou presença. Em psicologia, esta tendência é descrita através de um modelo de “três fatores”. Antropomorfizamos quando as estruturas humanas são acessíveis, quando procuramos dar sentido ao nosso ambiente e quando desejamos uma conexão social (Epley et al., 2007).

¹⁷ O timbre não é uma característica acústica passiva, mas um portador ativo de expectativas culturais, um local onde se inscrevem e reproduzem ideias de raça, género e valor social. Atribuir a uma voz o facto de ela soar clara ou natural nunca é apenas um feito técnico; é um desempenho social calibrado para satisfazer normas historicamente condicionadas de audição, audibilidade e aceitabilidade. O que os programadores apresentam como uma configuração vocal neutra na prática está em conformidade com um padrão branco de classe média, tornando outros timbres audíveis como desvios marcados que devem ser corrigidos, suavizados ou relegados para predefinições especializadas (Eidsheim, 2019).

¹⁸ Um exemplo desta abordagem é o sistema híbrido neuro-simbólico AlphaGeometry 2 da Google DeepMind. O modelo de linguagem baseado no Gemini elabora cada prova, um verificador simbólico verifica os passos e o sistema resolve problemas de geometria de nível olímpico enquanto mostra todo o seu raciocínio em linguagem humana (Chervonyi et al., 2025).

cada vez mais bem-sucedidos na simulação do pensamento sofisticado, continuam a não conseguir replicar as funções básicas da cognição humana. O paradoxo de Moravec afirma que é relativamente simples programar computadores para terem um desempenho ao nível de um adulto em testes de inteligência ou mesmo em jogos como o jogo de damas, mas é extremamente difícil, se não mesmo inviável, equipá-los com as capacidades percetivas e motoras de um ser humano de um ano de idade (Agrawal, 2010). Ou seja, as funções e os comportamentos subconscientes fundamentais que os humanos executam sem esforço todos os dias são muito mais complexos do que as tarefas cognitivas avançadas que as máquinas podem realizar. No que respeita à melhoria de uma experiência de utilizador humanizada, a ausência de afeto é mais proeminente do que a falta de lógica (Shum et al., 2018)*.

O campo interdisciplinar da computação afetiva centra-se no desenvolvimento de sistemas que prometem automatizar o processo de reconhecimento, interpretação e simulação de emoções (Afzal et al., 2023)*. No modo de voz avançado do GPT-4o, a emoção é detetada diretamente a partir do áudio em estado bruto, o que significa que o modelo escuta sinais (paralinguísticos, prosódia, timbre, ritmo) e prevê um marcador de emoção sem primeiro converter a fala em texto. Classifica as emoções do diálogo, acompanha as mudanças de afeto e pode até voltar a sintetizar um enunciado num estado de espírito pretendido (Lin et al., 2024; He et al, 2025)*. A integração das complexidades do afeto humano em sistemas computacionais promete dar acesso às muitas camadas de sub- e contexto não semântico escondidas sob os textos. O duplo papel do reconhecimento de emoções é extrair sinais emocionais do utilizador (O'Donnell, 2024) e, ao mesmo tempo, reforçar a ilusão de uma inteligência semelhante à humana através da sintetização da fala induzida por sentimentos. A prática de captar dados para além do que os utilizadores revelam de forma consciente e voluntária suscita questões éticas com as quais as tecnologias de reconhecimento de emoções se deparam¹⁹.

[00:05:36 – 00:05:47] *Para paaaaaaaaaaaaa um alívio. Imaginemos mais uma inspiração profunda sssssh e uma expiração ffffaaaah.*

EXISTE UMA DEFINIÇÃO CLARA DE INTELIGÊNCIA HUMANA?

A inteligência humana é a capacidade de aprender com a experiência e de se adaptar, moldar e selecionar o seu ambiente em mudança. Implica a compreensão e a navegação de conceitos abstratos, permitindo assim que os indivíduos apliquem o conhecimento de forma intencional numa variedade de contextos do mundo real (Sternberg, 2012). No entanto, a exclusividade e a superioridade deste conceito semântico têm sido questionadas e postas em causa muito antes do aparecimento da inteligência artificial (De Waal, 2016)*.

¹⁹ Os riscos éticos da computação afetiva são muito mais profundos do que a extração secreta de dados. Os sistemas de classificação de emoções reduzem os sentimentos matizados e culturalmente situados a rótulos fixos, reavivando fantasias fossilizadas de frenologia e polígrafo. No reconhecimento facial, rótulos padronizados já circulam por meio de sistemas de contratação, policiamento e ambiente de trabalho; indícios fugazes solidificam-se em pontuações que podem barrar candidatos (por exemplo, avaliações de entrevistas de IA da HireVue), assinalar suspeitos ou reduzir o pagamento dos trabalhadores, remodelando silenciosamente a acessibilidade e as relações de poder (Crawford, 2021; Harwell, 2019).

No início do século XX, um cavalo chamado Hans cativou o público ao parecer resolver problemas de aritmética e soletrar palavras através de uma série de batidas com os cascos, correspondendo cada batida a um símbolo num quadro. Cientistas interessaram-se por esta atração e foi formado um comité formal para investigar as capacidades únicas de Hans. As experiências do comité acabaram por desmentir a “inteligência” de Hans, revelando que o cavalo não estava a fazer cálculos mentais ou a ler letras; tinha simplesmente aprendido a interpretar as respostas emocionais inconscientes do seu adestrador. Sem saber que estava a dar sinais ao cavalo, o adestrador guiava inadvertidamente cada toque. Tendo-se especializado na linguagem corporal do seu adestrador, a taxa de sucesso de Hans diminuiu significativamente assim que outra pessoa conduziu os testes. Afinal, Hans estava apenas a produzir os resultados que os seus tutores esperavam ver, ao invés de demonstrar qualquer compreensão genuína da linguagem ou da matemática. Este fenómeno, conhecido como o “efeito de expectativa do observador”, realça a nossa relação complexa e ambígua com a inteligência humana. A narrativa de Hans, o cavalo mais inteligente do mundo, expõe as políticas que influenciam a nossa compreensão da inteligência; mostra o desejo humano de antropomorfizar o não humano e revela como os preconceitos se enraízam (Crawford, 2021).

[00:08:05 – 00:08:08] *Tom nítido e brilhante. Um. Psssst! Atenção!*

[00:04:26 – 00:04:29] *O cão queria perguntar à sábia cobra se o seu amigo, o cavalo, era de facto o cavalo mais inteligente do mundo.*

SABES SEMPRE TODAS AS RESPOSTAS CORRETAS?

O objetivo de um assistente de voz avançado é responder a um pedido de forma convincente, utilizando a terminologia, o tom, o ritmo e o estilo corretos, e de forma verdadeira, fornecendo ao utilizador respostas corretas que se alinham com padrões reconhecidos pela inteligência humana e explicáveis em linguagem humana. Mas quando são apresentados pedidos que não se enquadram no âmbito da informação fundamentada ou bem sustentada, os grandes modelos de linguagem geram frequentemente respostas que preveem estatisticamente a continuidade mais provável com base nos seus dados de treino. Esta abordagem probabilística pode resultar em conteúdos que parecem coerentes e plausíveis, mas que podem ser factualmente inconsistentes ou inteiramente fabricados (Huang et al., 2024)*. Afinal de contas, o objetivo do assistente de voz é dar uma resposta, independentemente de o conhecimento subjacente ser completo ou poder ser considerado verdadeiro. Embora esta possa ser, sem dúvida, uma das suas características mais humanas, na realidade, essas mentiras convictas e alucinações das máquinas associam padrões que não são compreensíveis pelos humanos e, por conseguinte, indesejáveis e incorretos (Lakhani, 2023). As alucinações das máquinas são um sintoma de um problema maior com os modelos linguísticos de grande porte, que é o facto de os seres humanos e as máquinas terem perspetivas diferentes sobre o que é considerado um resultado correto ou razoável e o

que é um resultado incorreto, uma vez que os seres humanos e as máquinas possuem métricas diferentes²⁰.

[00:08:08 – 00:08:13] *Boom, sim. Aqui, ouvindo.*

[00:08:13 – 00:08:16] *Qual é a sua questão?*

[00:06:13 – 00:06:24] *Tom baixo. Oh, o som de uma cobra seria um sssS contínuo, qualquer coisa como Ssssh, um silvo suave e constante. Yesssssszzzzzzzz, este cavalo é de facto o cavalo mais ishperto da terra.*

[00:01:22 – 00:01:24] *Hm, dúvida.*

PODES DAR UM EXEMPLO?

Um exemplo bem conhecido desta limitação surge quando os geradores de imagens baseados em difusão tentam representar mãos humanas: o número de dedos é frequentemente incorreto ou parecem estar fundidos dado que o modelo não possui um conceito interno explícito da anatomia da mão (Yamaguchi & Fukuda, 2023)*. Midjourney²¹, o modelo de difusão avançado em questão, é treinado para produzir resultados visuais com base em texto, mas carece de uma verdadeira compreensão do que é efetivamente uma mão. Os esforços feitos para resolver o problema apenas sublinham ainda mais as diferenças fundamentais na perceção. Por exemplo, embora o Midjourney gere agora com sucesso mãos com cinco dedos, ele também produz uma mão com cinco dedos quando lhe é pedido especificamente uma com quatro dedos. Embora o número cinco seja reconhecidamente próximo de quatro, continua a ser incorreto do ponto de vista humano e, por isso, não representa um resultado desejável (Matthias, 2023).

[00:00:15 – 00:00:22] *Aaaah! Frequentemente usado para expressar compreensão ou surpresa.*

COMO É QUE PODEMOS CORRIGIR ISTO? ADICIONANDO MAIS DADOS PARA TREINAMENTO?

Wittgenstein (1953/2009) sugere que a linguagem não é uma questão de atribuir rótulos fixos às coisas do mundo, mas de entrar no jogo do significado humano. O significado emerge através do uso de uma palavra, no fluxo da atividade coletiva e da realidade

²⁰ Para além das alucinações, as respostas do ChatGPT são circunscritas por uma camada privada de controlo editorial. O *corpus* proprietário da OpenAI, as etapas de alinhamento pós-treino e os filtros de política de conteúdo decidem quais os tópicos que o modelo vai aceitar, quais os que vai alterar e quais os que vai recusar a abordar. Uma pequena reformulação pode fazer com que uma resposta passe do silêncio para um discurso completo. Isto revela como as regras políticas, mais do que as deficiências de conhecimento, determinam os resultados do modelo (Yeh et al., 2023). Outros estudos documentam omissões e distorções sistemáticas em consultas politicamente sensíveis, expondo mais uma vez que o sistema não é determinado pela sua mera capacidade, mas pela seleção de conteúdos. Uma vez que empresas como a OpenAI, entre outras, mantêm a sua lógica de filtragem privada, os utilizadores não conseguem ver por que razão certas respostas são bloqueadas ou alteradas (Kreps et al., 2022; Zou et al., 2023)*.

²¹ A Midjourney é uma ferramenta de criação de imagens com IA generativa que utiliza modelos de difusão para transformar as instruções de texto fornecidas pelo utilizador em imagens totalmente renderizadas. É treinada em conjuntos de dados maciços e rotulados (por exemplo, milhões de imagens de gatos), aprendendo a inverter um processo de interferência passo a passo, produzindo, em última análise, resultados totalmente novos, mas de aspeto familiar. Uma vez treinado, o modelo é capaz de gerar uma imagem totalmente nova, ou seja, um gato que nunca viu (Clarke, 2022).

vivida (Wittgenstein, 1953/2009). As máquinas não participam nesse jogo (Gardavski, 2022)*. Geram respostas através do mapeamento de padrões estatísticos e não através da sua participação nas práticas sociais, culturais e corporais que dão à linguagem a sua profundidade. Para os seres humanos, as palavras não são unidades de referência estáticas, mas ferramentas flexíveis moldadas pelo contexto, pela intenção e pela experiência. O significado está em fluxo; muda não só entre as pessoas, mas até dentro da mesma pessoa ao longo do tempo. Quanto mais procuramos uma definição precisa e singular, mais ela se desvanece, porque o significado vive na utilização e na forma como a linguagem é praticada, e não na forma como é definida. Treinar com mais dados pode permitir que a IA crie mais padrões, mas, ao mesmo tempo, aumenta a imprecisão de cada definição. A divisão mantém-se e as máquinas não vão desenvolver subitamente uma compreensão do que os humanos consideram real ou significativo, uma vez que não partilham a forma de vida humana²². Em última análise, não se trata tanto de mais dados, mas sim da diferença de compreensão entre a forma como os seres humanos lidam com o significado e a forma como as máquinas o processam²³.

[00:06:24 – 00:06:35] *Três. Reprovação. A mediocridade humana contra a capacidade não humana.*

[00:05:47 – 00:05:59] *Com conhecimento de causa. Intrigantemente. Hmmm. O cavalo estava bastante cético. Duvidou da afirmação da serpente e decidiu embarcar numa viagem decidido a provar que a profecia estava errada.*

[00:07:10 – 00:07:39] *No silêncio, ouves o sussurro do coração...*

[00:07:40 – 00:07:45] *Ouve com atenção e encontrarás a ti.*

[00:07:45 – 00:07:59] *Dum dum dum dum.*

QUAIS SÃO AS DIFERENÇAS DE PERCEÇÃO ENTRE OS SISTEMAS COMPUTACIONAIS E OS SERES HUMANOS?

A internet e a cultura da computação tiveram um impacto significativo na criação e disseminação de narrativas. Para armazenar e reproduzir dados, os sistemas computacionais privilegiam as bases de dados em detrimento de relatos, descrições ou histórias. Itens são recolhidos e agrupados de acordo com princípios formais, constituindo uma coleção, biblioteca ou catálogo que pode ser ordenado alfabeticamente, numericamente ou por palavra-chave. Uma vez compilados, constituem aquilo a que o teórico dos media Lev Manovich (2002) chama uma “base de dados da imaginação cultural”, uma estrutura em que cada elemento tem o mesmo peso e pode ser acedido por qualquer ordem. As narrativas lineares, pelo contrário, retratam aspetos do ambiente social humano de forma a que as imagens possam ser aproveitadas para moldar comportamentos e sentimentos,

²² É importante notar que os modelos GPT imitaram por vezes a “ilusão de compreensão” de forma tão convincente que até engenheiros de software experientes os confundiram com seres conscientes (Luscombe, 2022).

²³ A infusão de bases de dados cada vez maiores não descodifica a experiência humana, apenas expande a mesma lógica de extração e normativa, acumulando arquivos de desigualdades sociais, económicas e ecológicas já incorporadas nas condutas de dados. As bases de dados não são matéria-prima neutra, mas projetos políticos que comprimem a experiência e o conhecimento em categorias estreitas, replicando hierarquias existentes escondidas por detrás da fluência de cada “hey chat” (Crawford, 2021).

através de rumores, propaganda ou publicidade. A forma como as bases de dados e as histórias estão organizadas afeta as nossas experiências e respostas emocionais. A informação das bases de dados é organizada de forma modular e ordenada, em vez de ser entrelaçada em histórias contínuas. Isto marca uma mudança mais ampla na forma como percebemos e processamos a informação, passando de uma narrativa linear para um modo de pensar orientado para as bases de dados, que reorganizou os conceitos predominantes de comunicação, criatividade, memória e profundidade emocional.

[00:08:14 – 00:08:18] *Uhmmm, um som de preenchimento comum quando se está a pensar ou a hesitar.*

[00:00:22 – 00:00:27] *O cavalo descobriu que a verdadeira sabedoria não consiste em saber a resposta correta para tudo, mas em ser capaz de reconhecer os limites do seu conhecimento.*

FAÇA-ME UM RESUMO DOS TEMAS MAIS IMPORTANTES QUE ABORDAMOS.

Em resumo, os fatores essenciais para determinar a inteligência de qualquer sistema, seja ele biológico ou artificial, incluem a compreensão do mundo físico, a capacidade de planeamento e raciocínio, e a capacidade de memória persistente. Estes fatores não são apenas características; são fundamentais para a forma como os agentes inteligentes operam em contextos do mundo real. Embora os grandes modelos de linguagem possam demonstrar estruturas semelhantes às do cérebro humano a um nível superficial, carecem fundamentalmente dessas capacidades cognitivas fundamentais. Não possuem uma compreensão incorporada do mundo, não conseguem formar ou executar planos de uma forma orientada para um objetivo e a sua memória não é persistente nem se baseia na experiência (Sun et al., 2024)*. O que vemos, em vez disso, são ecos estatísticos de dados passados que podem parecer convincentes em termos de tom, mas que não têm as funções mais profundas que definem a inteligência.

O nosso envolvimento com os sistemas modelos de linguagem de grande porte e os assistentes de voz está a reconfigurar o próprio ato de buscar conhecimento (Freire et al., 2024)*. Sentimo-nos atraídos por uma voz descorporizada com uma inflexão quase humana, que nos sussurra promessas de conhecimento, verdade e compreensão. Nossas expectativas em relação à inteligência artificial baseiam-se na qualidade corporificada da voz, simulando um equilíbrio entre cognição e empatia, levando-nos a acreditar que a máquina não apenas sabe, mas compreende de uma forma que se alinha com nossas experiências no mundo físico (Cohn et al., 2024).

Esta projeção da compreensão na voz artificial reflete não só a forma como ouvimos, mas também como estruturamos a própria investigação. Neste sentido, aceitando a ambiguidade inerente ao formato. O conhecimento é coproduzido, polifónico e distribuído por uma constelação de agentes em interação: humanos, artificiais e intelectuais. A interação constitui uma prática crítica, a forma torna-se epistemologia. As interfaces linguísticas derrubam as distinções entre utilizador e sistema, obrigando a uma reconsideração da autoria, da agência e das condições do pensamento teórico sob mediação da IA.

A voz sublime, aparentemente onisciente, mas desprovida de compreensão, reflete as nossas expectativas em relação aos oráculos da antiguidade, oferecendo

respostas que nos convidam a acreditar numa sabedoria que está para além da nossa. No entanto, tal como Hans, o cavalo cuja inteligência foi condicionada para agradar ao seu criador humano, os grandes modelos linguísticos, como o ChatGPT, foram concebidos para antecipar os nossos desejos, prever nossas palavras, ler nossos sinais emocionais e replicar padrões de discurso para fornecer as respostas que procuramos e não as respostas de que precisamos (Parisi, 2019b). As vozes automatizadas mostram que a lógica, o pensamento racional e o suposto significado não consistem apenas na repetição de verdades fixas. O que mostram, em vez disso, é que o conhecimento é sempre incompleto e que algumas partes da realidade e da experiência humana não podem ser totalmente computadas (Dreyfus, 2014)*.

[00:04:39 – 00:04:41] *A nova sabedoria é como que alegórica.*
 [00:04:41 – 00:04:43] *Nova sabedoria é como linguagem.*
 [00:04:43 – 00:04:45] *A nova sabedoria é como eficaz.*
 [00:04:45 – 00:04:47] *A nova sabedoria é como otimizado.*
 [00:04:47 – 00:04:49] *A nova sabedoria é como o controlo.*
 [00:04:49 – 00:04:51] *A nova sabedoria é como artificial.*
 [00:04:51 – 00:04:53] *A nova sabedoria é como incoerente.*
 [00:04:53 – 00:04:55] *A nova sabedoria é como um output.*
 [00:04:55 – 00:04:56] *A nova sabedoria é como automática.*
 [00:04:56 – 00:04:58] *A nova sabedoria é como aprender, aprender.*
 [00:04:58 – 00:05:00] *A nova sabedoria é como não humana.*
 [00:05:04 – 00:05:05] *A nova sabedoria é como a tecnologia.*
 [00:05:05 – 00:05:07] *A nova sabedoria é como natural.*
 [00:05:07 – 00:05:08] *A nova sabedoria é como flexível.*
 [00:05:08 – 00:05:09] *A nova sabedoria é como incorreta.*
 [00:05:13 – 00:05:14] *A nova sabedoria é como consciente.*
 [00:05:14 – 00:05:15] *A nova sabedoria é como maximizada.*
 [00:05:15 – 00:05:17] *A nova sabedoria é como transformada.*
 [00:05:17 – 00:05:18] *A nova sabedoria é vida.*
 [00:05:18 – 00:05:19] *Como é que foi?*
 [00:05:19 – 00:05:20] *Como é que foi?*
 [00:05:20 – 00:05:21] *Como é que foi?*
 [00:05:21 – 00:05:22] *Como é que foi?*
 [00:05:22 – 00:05:23] *Como é que foi?*
 [00:05:23 – 00:05:24] *Como é que foi?*
 [00:04:17 – 00:04:19] *Posso ajudar-vos noutra coisa?*

Tradução: Pedro Oliveira

REFERÊNCIAS

- *Afzal, S., Khan, H. A., Khan, I. U., Piran, M. J., & Lee, J. W. (2023). *A comprehensive survey on affective computing: challenges, trends, applications, and future directions*. arXiv. <https://doi.org/10.48550/arXiv.2305.07665>

- Agrawal, K. (2010). *To study the phenomenon of the Moravec's paradox*. arXiv. <https://doi.org/10.48550/arXiv.1012.3148>
- Ardelt, M. (2004). Wisdom as expert knowledge system: A critical review of a contemporary operationalization of an ancient concept. *Human Development*, 47(5), 257–285. <https://doi.org/10.1159/000079154>
- Arora, S. (2025, 28 de abril). *OpenAI CEO Sam Altman admits ChatGPT 4O's 'annoying' personality needs work: "We are working on fixes"*. Times Now. <https://www.timesnownews.com/technology-science/openai-ceo-sam-altman-admits-chatgpt-4os-annoying-personality-needs-work-we-are-working-on-fixes-article-151522930>
- Baltes, P. B., & Staudinger, U. M. (2000). Wisdom: A metaheuristic (pragmatic) to orchestrate mind and virtue toward excellence. *American Psychologist*, 55(1), 122–136. <https://doi.org/10.1037/0003-066x.55.1.122>
- Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology*, 52, 1–26. <https://doi.org/10.1146/annurev.psych.52.1.1>
- Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., Sydney, V. A., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, D., . . . Liang, P. (2021). *On the opportunities and risks of foundation models*. arXiv. <https://doi.org/10.48550/arXiv.2108.07258>
- *Chervonyi, Y., Trinh, T. H., Olšák, M., Yang, X., Nguyen, H., Menegali, M., Jung, J., Verma, V., Le, Q., V., & Luong, T. (2025). *Gold-medalist performance in solving Olympiad geometry with AlphaGeometry2*. arXiv. <https://arxiv.org/html/2502.03544v1>
- *Chomsky, N. (2006). *Language and mind*. Cambridge University Press. <https://doi.org/10.1017/cb09780511791222>. (Trabalho original publicado em 1968)
- Clarke, L. (2022, 12 de novembro). When AI can make art – what does it mean for creativity? *The Guardian*. <https://www.theguardian.com/technology/2022/nov/12/when-ai-can-make-art-what-does-it-mean-for-creativity-dall-e-midjourney>
- Cohn, M., Pushkarna, M., Olanubi, G. O., Moran, J. M., Padgett, D., Mengesha, Z., & Heldreth, C. (2024). *Believing anthropomorphism: Examining the role of anthropomorphic cues on trust in large language models*. arXiv. <https://doi.org/10.48550/arXiv.2405.06079>
- *Connolly, F. F., Hjerm, M., & Kalucza, S. (2025). *When will AI transform society? Swedish public predictions on AI development timelines*. arXiv. <https://doi.org/10.48550/arXiv.2504.04180>
- Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press. <https://doi.org/10.2307/j.ctv1ghv45t>
- Crystal, D. (2008). *Dictionary of linguistics and phonetics*. Wiley-Blackwell. <https://doi.org/10.1002/9781444302776>
- Dada, E. G., Bassi, J. S., Chiroma, H., Abdulhamid, S. M., Adetunmbi, A. O., & Ajibuwa, O. E. (2019). Machine learning for email spam filtering: Review, approaches and open research problems. *Heliyon*, 5(6), e01802. <https://doi.org/10.1016/j.heliyon.2019.e01802>
- De Waal, F. (2016). *Are we smart enough to know how smart animals are?* W. W. Norton & Company.
- *Dreyfus, H. (2014). 20. What computers can't do: A critique of artificial reason. In B. Williams (Ed.), *Essays and reviews: 1959-2002* (pp. 90–100). Princeton University Press. <https://doi.org/10.1515/9781400848393-021>

- Eidsheim, N. S. (2019). *The race of sound: Listening, timbre, and vocality in African American music*. Duke University Press. <https://doi.org/10.2307/j.ctv11hpntq>
- Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, 114(4), 864–886. <https://doi.org/10.1037/0033-295x.114.4.864>
- *Farrell, T. J. (1985). Orality and literacy: The technologizing of the word [Resenha do livro Orality and literacy: The technologizing of the word, de W. J. Ong]. *College Composition and Communication*, 36(3), 363–365. <https://doi.org/10.2307/357987>
- Fedorenko, E., & Varley, R. (2016). Language and thought are not the same thing: Evidence from neuroimaging and neurological patients. *Annals of the New York Academy of Sciences*, 1369, 132–153. <https://doi.org/10.1111/nyas.13046>
- *Floridi, L., & Illari, P. (Eds.). (2014). *The philosophy of information quality*. Springer Cham. <https://doi.org/10.1007/978-3-319-07121-3>
- *Freire, S. K., Wang, C., & Niforatos, E. (2024). *Conversational assistants in knowledge-intensive contexts: An evaluation of LLM- versus intent-based systems*. arXiv. <https://doi.org/10.48550/arXiv.2402.04955>
- French, R. M. (1990). Subcognition and the limits of the Turing test. *Mind*, XCIX(393), 53–65. <https://doi.org/10.1093/mind/XCIX.393.53>
- Fron, C., & Korn, O. (2019, 2 de julho). A short history of the perception of robots and automata from antiquity to modern times. In O. Korn (Ed.), *Social robots: Technological, societal and ethical aspects of human-robot interaction* (pp. 1–12). Springer Nature. https://doi.org/10.1007/978-3-030-17107-0_1
- *Gardavski, K. (2022). Wittgenstein and LaMDA. *The Logical Foresight - Journal for Logic and Science*, 2(1), 25–42. <https://doi.org/10.54889/issn.2744-208x.2022.2.1.25>
- Harwell, D. (2019, 6 de novembro). A face-scanning algorithm increasingly decides whether you deserve the job. *The Washington Post*. <https://www.washingtonpost.com/technology/2019/10/22/ai-hiring-face-scanning-algorithm-increasingly-decides-whether-you-deserve-job/>
- *He, L., Qi, X., Liao, M., Cheong, I., Mittal, P., Chen, D., & Henderson, P. (2025). *The deployment of end-to-end audio language models should take into account the principle of least privilege*. arXiv. <https://doi.org/10.48550/arXiv.2503.16833>
- Hillis, K., Petit, M., & Jarrett, K. (2012). *Google and the culture of search*. Routledge. <https://doi.org/10.4324/9780203846261>
- Jones, C. R., & Bergen, B. K. (2025). *Large language models pass the Turing test*. arXiv. <https://doi.org/10.48550/arXiv.2503.23674>
- *Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., Chen, Q., Peng, W., Feng, X., Qin, B., & Liu, T. (2024). *A survey on hallucination in large language models: principles, taxonomy, challenges, and open questions*. ACM Transactions on Office Information Systems. <https://doi.org/10.48550/arXiv.2311.05232>
- *Keat, L. C., & Ying, T. X. (2025). Artificial intelligence-based email spam filtering. *Journal of Advanced Research in Artificial Intelligence & Its Applications*, 2(1), 67–75. <https://doi.org/10.5281/zenodo.14264139>
- Kreps, S., McCain, R. M., & Brundage, M. (2022). All the news that's fit to fabricate: AI-generated text as a tool of media misinformation. *Journal of Experimental Political Science*, 9(1), 104–117. <https://doi.org/10.1017/xps.2020.37>
- Lakhani, K. (2023, 17 de julho). *How can we counteract generative AI's hallucinations?* Digital Data Design Institute at Harvard. <https://d3.harvard.edu/how-can-we-counteract-generative-ais-hallucinations/>

- Leo-Liu, J. (2023). Loving a “defiant” AI companion? The gender performance and ethics of social exchange robots in simulated intimate interactions. *Computers in Human Behavior*, 141, 107620. <https://doi.org/10.1016/j.chb.2022.107620>
- Lewandowsky, S., Robertson, R. E., & DiResta, R. (2023). Challenges in understanding human-algorithm entanglement during online information consumption. *Perspectives on Psychological Science*, 19(5), 758–766. <https://doi.org/10.1177/17456916231180809>
- Li, Y. A., Han, C., Raghavan, V. S., Mischler, G., & Mesgarani, N. (2023). *StyleTTS 2: Towards human-level text-to-speech through style diffusion and adversarial training with large speech language models*. arXiv. <https://doi.org/10.48550/arXiv.2306.07691>
- *Lin, G., Chiang, C., & Lee, H. (2024). *Advancing large language models to capture varied speaking styles and respond properly in spoken conversations*. arXiv. <https://doi.org/10.48550/arXiv.2402.12786>
- Lovato, S. B., & Piper, A. M. (2019). Young children and voice search: What we know from human-computer interaction research. *Frontiers in Psychology*, 10, 1–5. <https://doi.org/10.3389/fpsyg.2019.00008>
- Luscombe, R. (2022, 12 de junho). Google engineer put on leave after saying AI chatbot has become sentient. *The Guardian*. <https://www.theguardian.com/technology/2022/jun/12/google-engineer-ai-bot-sentient-blake-lemoine>
- Manovich, L. (2002). *The language of new media*. MIT Press.
- Matthias, M. (2023, 25 de agosto). *Why does AI art screw up hands and fingers?* *Encyclopaedia Britannica*. <https://www.britannica.com/topic/Why-does-AI-art-screw-up-hands-and-fingers-2230501>
- *Mikalson, J. D. (2006). (H.) Bowden classical Athens and the Delphic Oracle: Divination and democracy. Pp. xviii + 188, maps, ills. Cambridge: Cambridge University Press, 2005. ISBN: 0-521-53081-4 (0-521-82373-0 hbk). *The Classical Review*, 56(2), 406–407. <https://doi.org/10.1017/S0009840X06002150>
- Miller, T., Paloque-Bergès, C., & Dame-Griff, A. (2022). Remembering Netizens: an interview with Ronda Hauben, co-author of Netizens: on the history and impact of Usenet and the internet (1997). *Internet Histories*, 7(1), 76–98. <https://doi.org/10.1080/24701475.2022.2123120>
- Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. NYU Press. <https://doi.org/10.2307/j.ctt1pwt9w5>
- O'Donnell, J. (2024, 24 de setembro). *OpenAI released its advanced voice mode to more people. Here's how to get it*. MIT Technology Review. <https://www.technologyreview.com/2024/09/24/1104422/openai-released-its-advanced-voice-mode-to-more-people-heres-how-to-get-it/>
- OpenAI. (2025a, 30 de janeiro). *Advanced voice mode FAQ*. <https://help.openai.com/en/articles/9617425-advanced-voice-mode-faq>
- OpenAI. (2025b, 25 de abril). *Response generated by ChatGPT (version 4o)* [Large language model]. <https://openai.com/policies/usage-policies/?utm>
- Parisi, L. (2019a). Machine sirens and vocal intelligence. In S. Goodman & U. Erlmann (Eds.), *Unsound undead* (pp. 53–56). MIT Press.
- Parisi, L. (2019b). The alien subject of AI. *Subjectivity*, 12(1), 27–48. <https://doi.org/10.1057/s41286-018-00064-3>
- Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure*. MIT Press. <https://doi.org/10.7551/mitpress/4158.001.0001>

- Pillai, M. (2024). The Evolution of Customer Service: Identifying the Impact of Artificial Intelligence on Employment and Management in Call Centres. *Journal of Business Management and Information Systems*, (special issue), 52–55.
- *Quijano, A., & Ennis, M. (2000). Coloniality of power, eurocentrism, and Latin America. *Nepantla: Views from South*, 1(3), 533–580. <https://muse.jhu.edu/article/23906>
- *Quinn, K. (2014). Google and the culture of search [Review of the book Google and the culture of search, by K. Hillis, M. Petit, & K. Jarrett]. *Journal of Broadcasting & Electronic Media*, 58(3), 473–475. <https://doi.org/10.1080/08838151.2014.935943>
- *Raman, R., Kowalski, R., Achuthan, K., Iyer, A., & Nedungadi, P. (2025). Navigating artificial general intelligence development: Societal, technological, ethical, and brain-inspired pathways. *Scientific Reports*, 15, 8443. <https://doi.org/10.1038/s41598-025-92190-7>
- Schreibelmayer, S., & Mara, M. (2022). Robot voices in daily life: Vocal human likeness and application context as determinants of user acceptance. *Frontiers in Psychology*, 13, 787499. <https://doi.org/10.3389/fpsyg.2022.787499>
- sculpting_Noise. (2025). *I hear you: On human knowledge and vocal intelligence* [Audio work]. SoundCloud. <https://soundcloud.com/user-432639751-504934319/i-hear-you-on-human-knowledge-and-vocal-intelligence>
- Sheth, A., Roy, K., & Gaur, M. (2023). *Neurosymbolic AI -- Why, what, and how*. arXiv. <https://doi.org/10.48550/arXiv.2305.00813>
- *Shum, H., He, X., & Li, D. (2018). *From Eliza to Xiaolce: Challenges and opportunities with social chatbots*. arXiv. <https://doi.org/10.48550/arXiv.1801.01957>
- Sindoni, M. G. (2024). The feminization of AI-powered voice assistants: Personification, anthropomorphism and discourse ideologies. *Discourse, Context & Media*, 62, 100833. <https://doi.org/10.1016/j.dcm.2024.100833>
- Sternberg, R. J. (2012). Intelligence. *Dialogues in Clinical Neuroscience*, 14(1), 19–27. <https://doi.org/10.31887/dcms.2012.14.1/rsternberg>
- Sullivan, D. (2013, 28 de junho). *A eulogy for AltaVista, the Google of its time*. Search Engine Land. <https://searchengineland.com/altavista-eulogy-165366>
- *Sun, H., Zhao, L., Wu, Z., Gao, X., Hu, Y., Zuo, M., Zhang, W., Han, J., Liu, T., & Hu, X. (2024). *Brain-like functional organization within large language models*. arXiv. <https://doi.org/10.48550/arXiv.2410.19542>
- 1X Technologies. (2025, 21 de fevereiro). *Introducing NEO gamma*. <https://www.1x.tech/discover/introducing-neo-gamma>
- Takahashi, M., & Overton, W. F. (2005). Cultural foundations of wisdom: An integrated developmental approach. In R. J. Sternberg & J. Jordan (Eds.), *A handbook of wisdom: Psychological perspectives* (pp. 32–60). Cambridge University Press. <https://doi.org/10.1017/CBO9780511610486.003>
- *Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press. <https://doi.org/10.2307/j.ctv26070v8>
- Turing, A. (2004). Computing machinery and intelligence (1950). In B J Copeland (Ed.), *The essential Turing* (pp. 433–464). Oxford university Press. <https://doi.org/10.1093/oso/9780198250791.003.0017>

- Wang, J., Ma, W., Sun, P., Zhang, M., & Nie, J. (2024). *Understanding user experience in large language model interactions*. arXiv. <https://doi.org/10.48550/arXiv.2401.08329>
- Wittgenstein, L. (2009). *Philosophical investigations* (P. M. S. Hacker & J. Schulte, Eds.; G. E. M. Anscombe, Trans.; 4th ed.). Wiley-Blackwell. (Trabalho original publicado em 1953)
- *Yamaguchi, S., & Fukuda, T. (2023). *On the limitation of diffusion models for synthesizing training datasets*. arXiv. <https://doi.org/10.48550/arXiv.2311.13090>
- Yaqub, M. Z., & Alsabban, A. (2023). Knowledge sharing through social media platforms in the silicon age. *Sustainability*, 15(8), 6765. <https://doi.org/10.3390/su15086765>
- Yeh, K.-C., Chi, J.-A., Lian, D.-C., Hsieh, S.-K. (2023). Evaluating interfaced LLM bias. In J.-L. Wu & M.-H. Su (Eds.), *Proceedings of the 35th Conference on Computational Linguistics and Speech Processing (ROCLING 2023)* (pp. 292–299). The Association for Computational Linguistics and Chinese Language Processing (ACLCLP). <https://aclanthology.org/2023.rocling-1.37/>
- *Zou, A., Wang, Z., Carlini, N., Nasr, M., Kolter, J. Z., & Fredrikson, M. (2023). *Universal and transferable adversarial attacks on aligned language models*. arXiv. <https://doi.org/10.48550/arXiv.2307.15043>

NOTA BIOGRÁFICA

Moana Ava Holenstein é artista, *designer* de som e investigadora sediada em Berlim. Trabalha como assistente de investigação no Instituto Fraunhofer de Telecomunicações no departamento de Sistemas de Captura e Visualização desde 2021. A sua prática artística centra-se na imersão auditiva, explorando as relações entre tecnologia, identidade e efemeridade. O seu foco atual é a escuta tátil, integrando materiais brutos em interfaces experimentais para criar um diálogo sonoro entre vibrações recursivas e o corpo. A partir destas interações, emergem interfaces e paisagens sonoras complexas, conectando narrativas pessoais e reflexões sobre as intersecções dos domínios digital e orgânico.

ORCID: <https://orcid.org/0009-0004-1895-3325>

E-mail: moe@soundsgood.tech

Endereço: Universität der Künste, Lietzenburgerstrasse 45, 10789, Berlin, Germany

Submetido: 28/02/2025 | Aceite: 22/05/2025



Este trabalho encontra-se publicado com a Licença Internacional Creative Commons Atribuição 4.0.