



Universidade Federal do ABC

Sobre polvos, papagaios e espelhos: IA generativa e as ilusões de sentido e pensamento

Cryptorave 2026, SP, 9/5/26

Prof. Miguel Said Vieira

miguel.vieira@ufabc.edu.br



Resumo

- como as tecnologias nos afetam
- LLMs:
 - papagaios estocásticos,
polvos enroladores (Bender)
 - espelhos ilusórios (Vallor)
- como afetar as tecnologias?

Como as tecnologias nos afetam?

- tecnologias não são neutras (“meios”);
olho vivo: que efeitos produzem?
 - que sujeitos constituem?
 - quem ganha / perde?
 - que mundos ajudam a construir / destruir?


Emily Bender: polvos e papagaios

- linguista computacional
- dois artigos sobre riscos e limitações de LLMs (IA generativa textual)
 - anteriores a ChatGPT!



Foto: Ian Allen ([New York Magazine](#))

Papagaios estocásticos: o artigo

- “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? ” (2021)
 - Gebru, Mitchell (2 das co-autoras):
ex-Google, equipe “IA Ética”
- revisão bibliográfica: riscos de LLMs
 - custos (e injustiças) ambientais
 - dados de treinamento problemáticos (inescrutáveis, enviesados...)
 - “papagaio estocástico”: argumento linguístico...

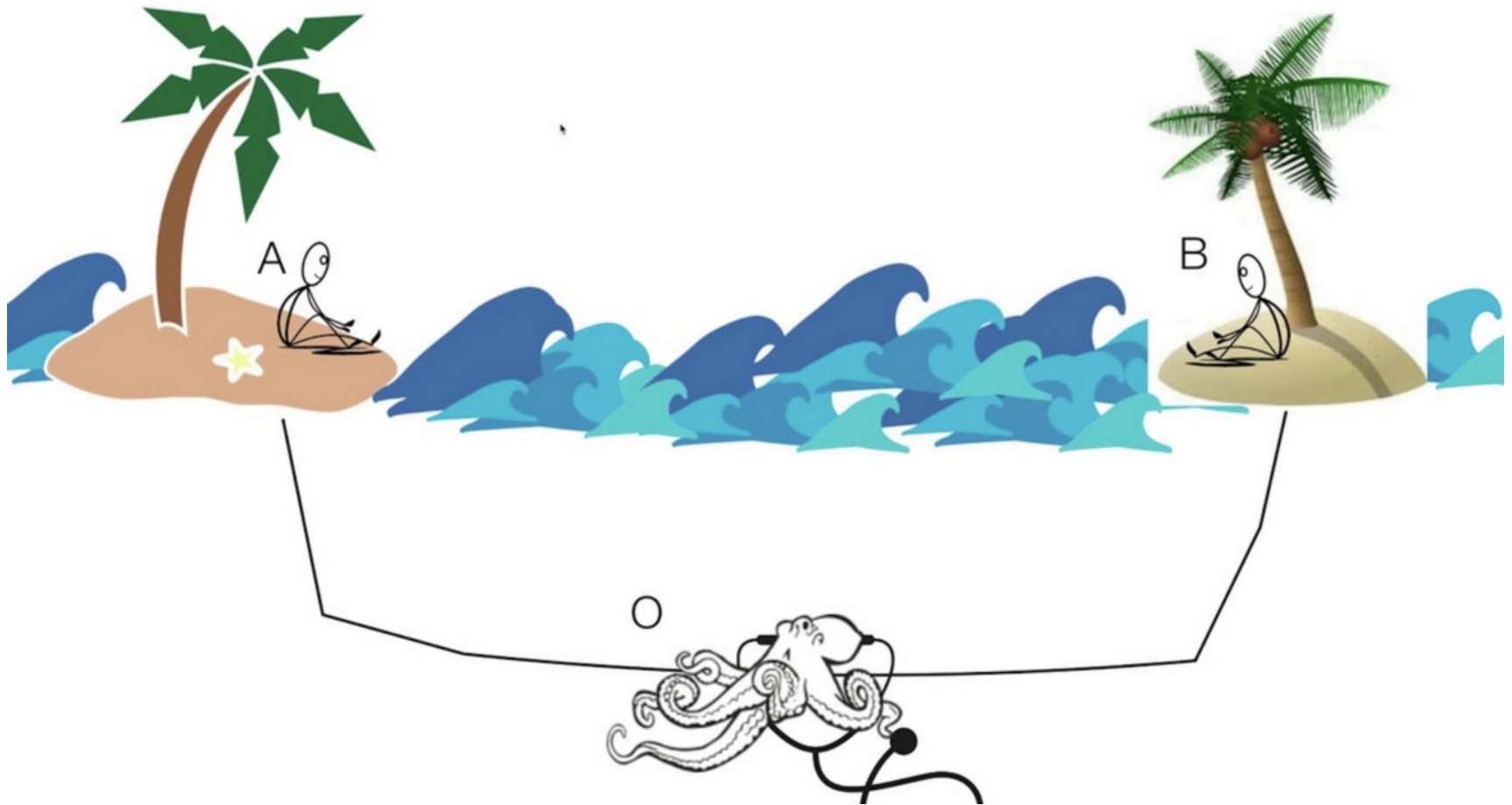
Papagaios estocásticos: a definição

- “[um LLM é] um sistema para costurar atabalhoadamente sequências de formas linguísticas observadas em seus muitos dados de treinamento, de acordo com informações probabilísticas sobre como elas são combinadas, mas sem nenhuma referência a sentido: um papagaio estocástico” [p. 617, trad. minha]

Polvos enroladores: o artigo

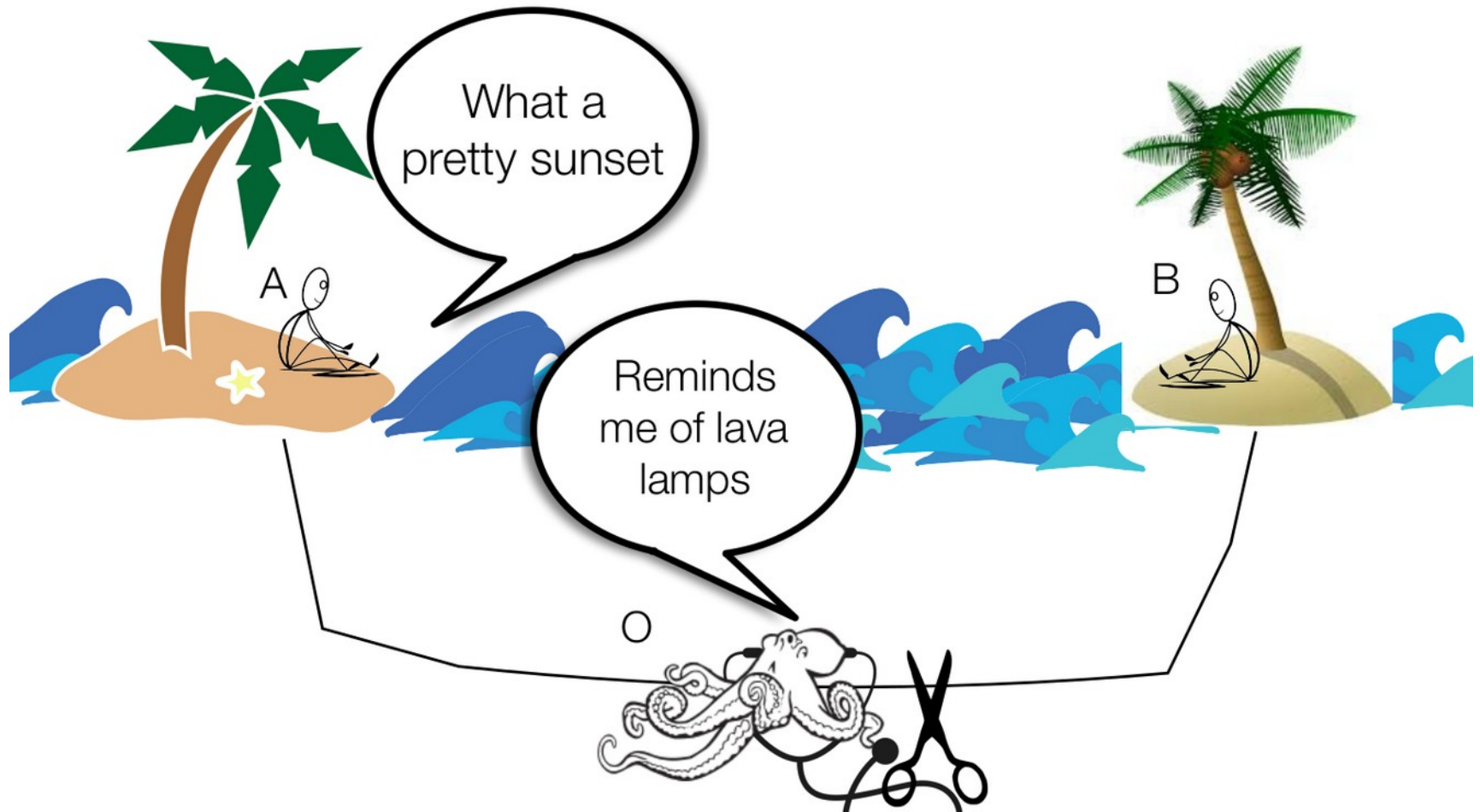
- “Climbing towards NLU” (2020)
 - formulação anterior (e mais extensa) desse argumento linguístico
 - ideia central:
 - como LLMs são treinadas apenas em *formas*, não há como aprenderem sobre *sentido*
 - *sentido*: relação entre uma forma linguística e uma intenção comunicativa (externa à linguagem)
 - » relação mais complexa que parece...

Polvos enroladores: o experimento mental

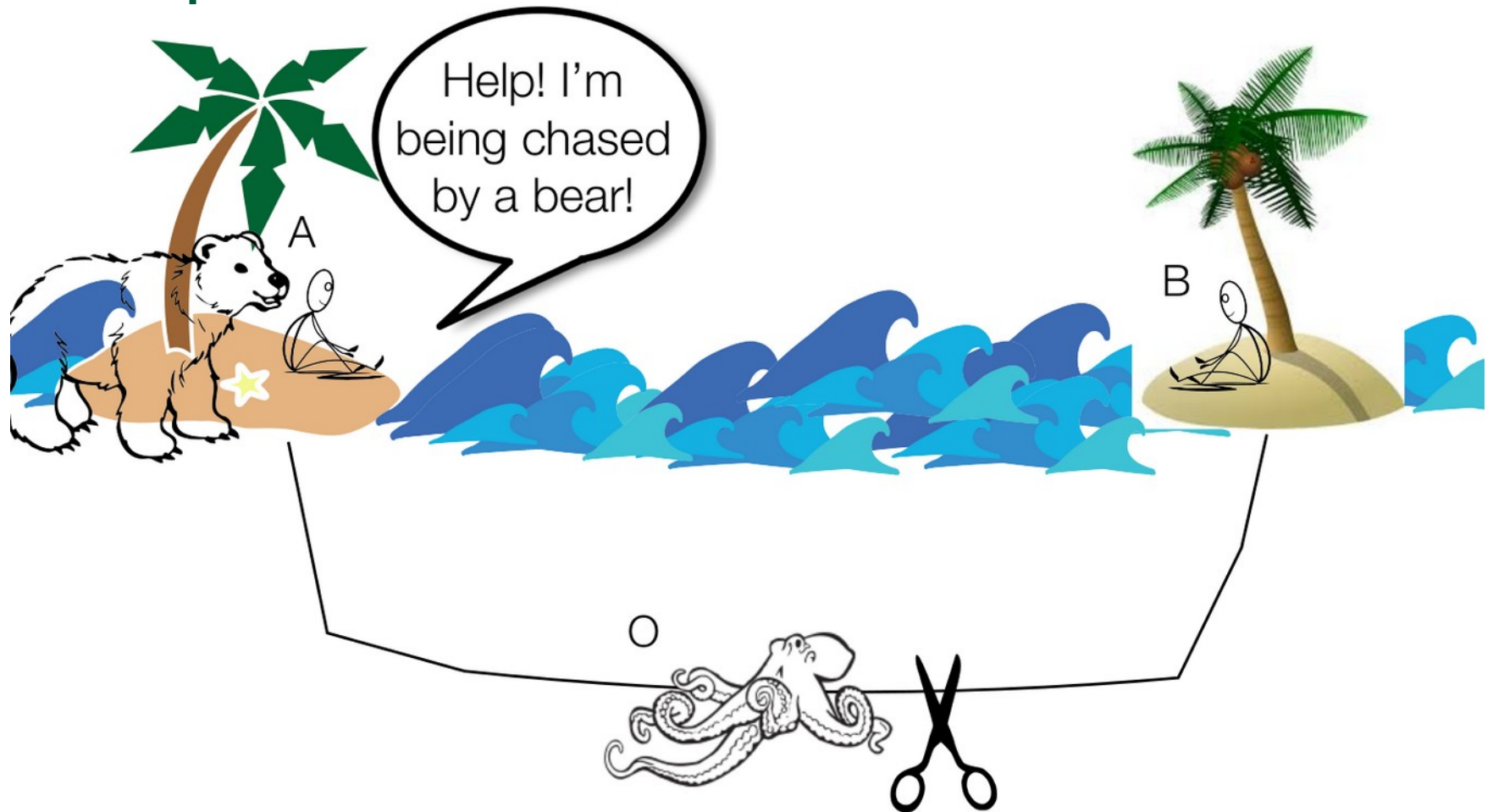


Fonte: [slides](#) de Bender & Koller

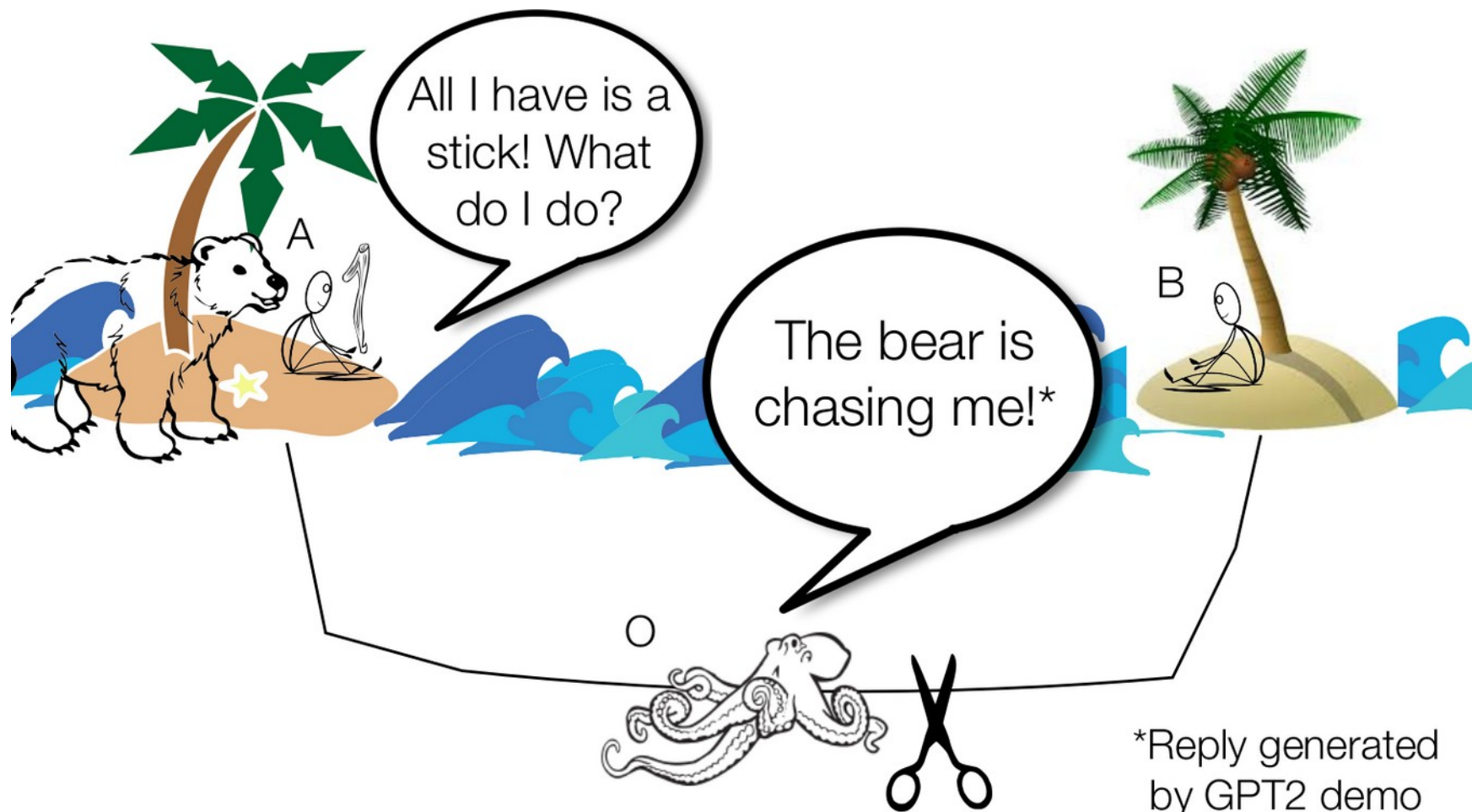
Polvos enroladores: o experimento mental



Polvos enroladores: o experimento mental



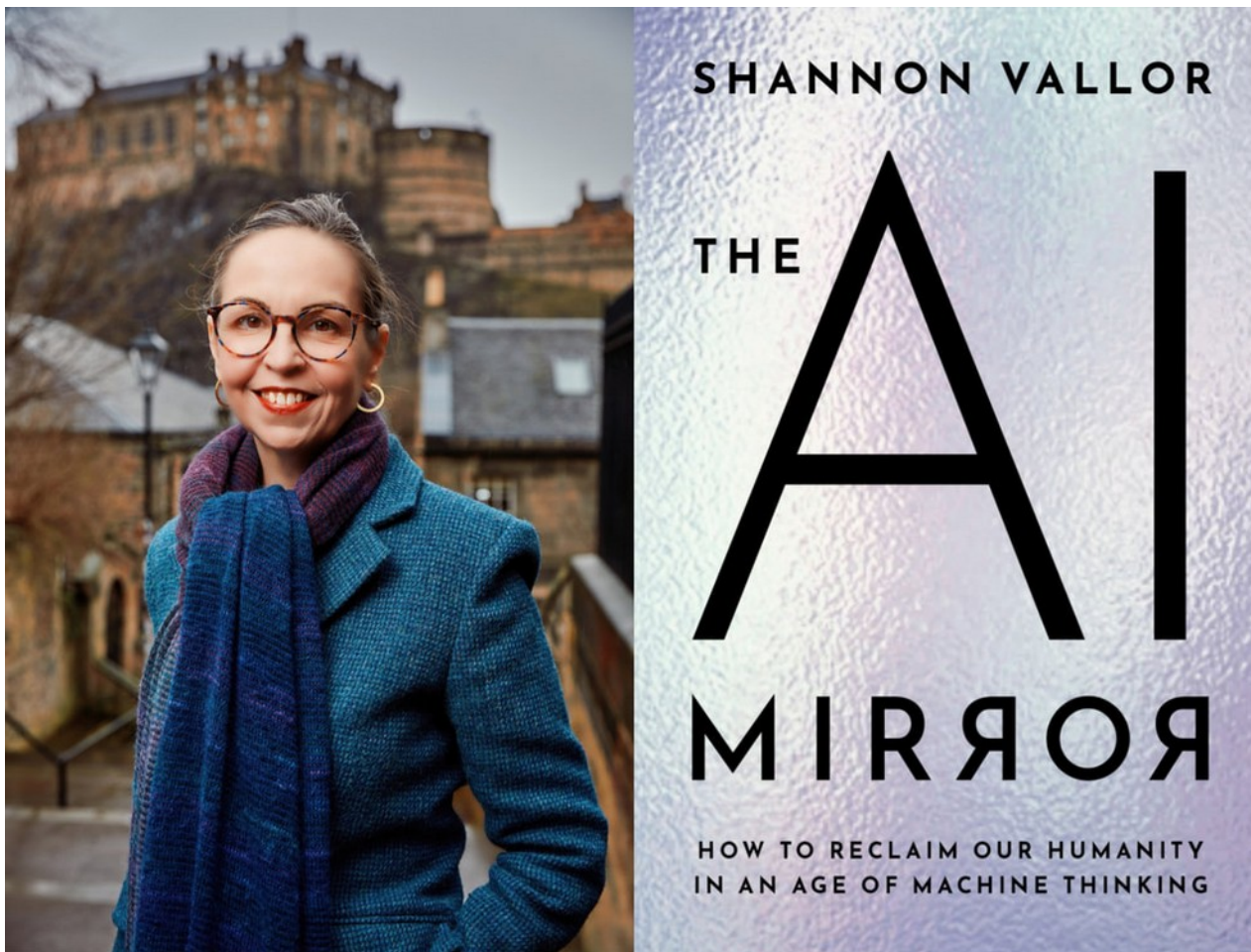
Quando a falta de sentido fica evidente



Importante: polvos enroladores dependem de escuta ativa

- as frases do polvo não possuíam *sentido*:
“A” é que encontrava sentido nelas
 - como as formas eram similares, acreditou que o polvo era “B”
 - para buscar sentido, assumiu intenção comunicativa do outro
 - ...como sempre fazemos ao dialogar!

Vallor: IA como espelho



Espelhos ilusórios

- LLMs como espelhos
 - reflexão (parcial, distorcida) de nossa cultura
 - vemos neles a fala de um ser humano
(e imaginamos pensamento, mente, pessoa)

Narciso e a reflexão: quem (se) ilude?



Narciso, por Caravaggio

Os problemas do espelho

- reflexo distorcido (viés), mas não só:
 - quando confiamos nele para julgar, interpretar, escolher, provoca amnésia de nossa agência e autoconhecimento
 - usamos para entender o presente e prever o futuro, mas reflete apenas o passado
 - o passado nos trouxe à polícrise do presente;
extrapolações do passado não a solucionarão

Os problemas do espelho

- simulacro da criatividade
 - criatividade não é repetição com variações; é expressão de um sujeito (que tem algo a dizer, e que se reinventa ao fazê-lo)
- “enrolação”
 - não tem concepção de verdade / mentira: só *enrola* (bem, e sem vergonha)
 - como filtrar as besteiras em pele de cordeiro?

E nós, podemos afetar as tecnologias?
Como?
(É com vocês...)

Como afetar as tecnologias (para inventar o futuro)

- sociedade também determina a tecnologia
 - IA atual materializa valores do Vale do Silício: eficiência, velocidade, fluidez, controle...
- tecnologia é produzida pela sociedade: não está acima dela, nem é “inevitável”
 - não dá pra “desinventar”, mas dá pra resistir, regular
- Vallor: ameaça não é o espelho (ou IAG), mas a ilusão que nutrimos com ele

Referências

- BENDER, Emily M. et al. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜. FAccT '21, Nova York: Association for Computing Machinery, 1 mar. 2021. p. 610–623. Disponível em: <<https://dl.acm.org/doi/10.1145/3442188.3445922>>.
- BENDER, Emily M.; KOLLER, Alexander. Climbing towards NLU: On meaning, form, and understanding in the age of data. 58th annual meeting of the association for computational linguistics. 2020, pp. 5185–5198. Disponível em: <<https://aclanthology.org/2020.acl-main.463/>>. Acesso em: 22 nov. 2024.
- VALLOR, Shannon. *The AI mirror: how to reclaim our humanity in the age of machine thinking*. Nova York: Oxford University Press, 2024.

Obrigado!

miguel.vieira@ufabc.edu.br

<https://miguelvieira.prof.ufabc.edu.br>