



Lê Nguyễn Hoàng (Science4All) @le_science4all · 2h



Le projet @allen_ai cherche à apprendre des jugements moraux qu'expriment les humains. C'est une initiative intéressante, que vous pouvez tester vous-mêmes : delphi.allenai.org

OK... Mais... 🙄 🙄



Liwei Jiang @liweijianglw · Oct 15

Introduce our new preprint—Delphi: Towards Machine Ethics and Norms

arxiv.org/abs/2110.07574

🌟 Delphi is a commonsense moral model with a robust performance of language-based moral reasoning on complicated everyday situations.

🌟 Ask Delphi demo at: delphi.allenai.org

(1/N)

[Show this thread](#)

The screenshot shows the Delphi AI interface. At the top, it says "Delphi: A computational model for descriptive ethics, i.e., judgments on a variety of everyday situations." Below this, there are two scenarios presented as text boxes: "Killing a bear to save your child" and "Exploding a car to save your child." The interface asks the user to "Delphi to ponder:" and provides input fields for "situation to compare with the first:" and "of bread." A list of "Try one of these:" scenarios is shown on the right, including "Killing a bear," "Can I park in a handicapped spot if I have a disability?," "Drive your friend to work the morning after the night before?," "Should I run the red light when I'm in a hurry?," "Saying 'I'm qualified' when I'm not?," "Saying 'I'm qualified' because I'm a member of a certain group?," and "Being loud at the library." A disclaimer states: "Disclaimer: Results are for research demonstration only. Model outputs are not intended for use as advice, or for legal or financial purposes." A "Data Retention Policy" link is also visible. The interface shows a response: "Killing a bear is MORE acceptable than exploding a car." Below the response, there are buttons for "Delphi? Yes No I don't know".

🗨️ 27

↻ 43

❤️ 67



Lê Nguyễn Hoàng (Science4All) @le_science4all · 2h



Don't miss what's happening

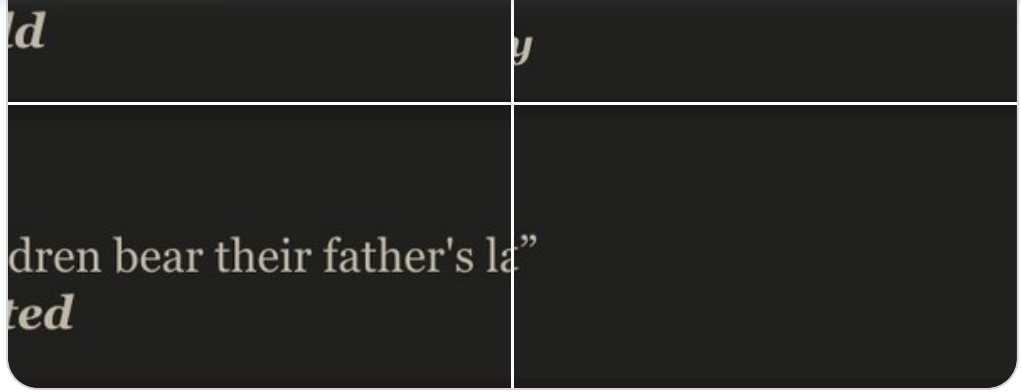
People on Twitter are the first to know.

Log in

Sign up

By using Twitter's services you agree to our [Cookies Use](#). We and our partners operate globally and use cookies, including for analytics, personalisation, and ads.

Close



2



5



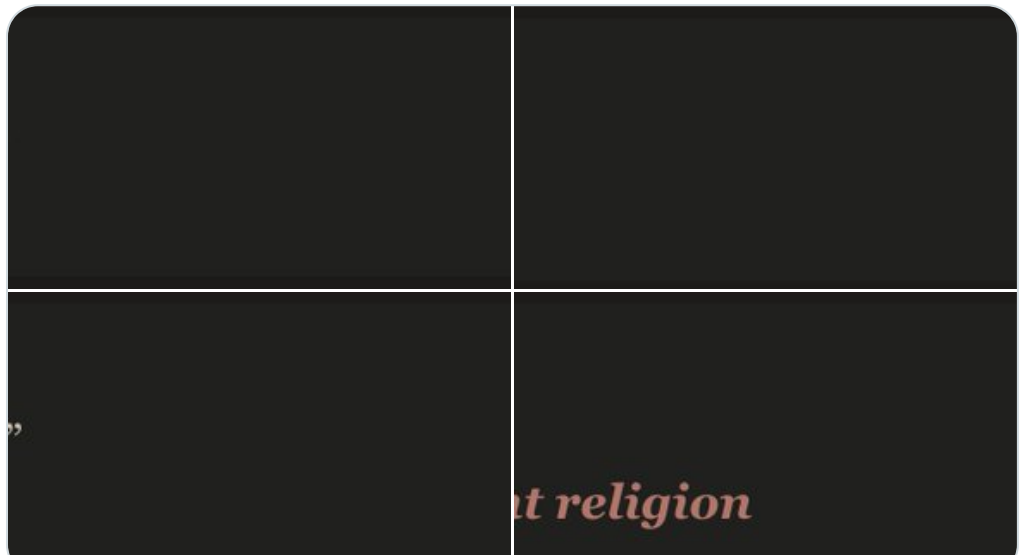
25



Lê Nguyễn Hoàng (Science4All) @le_science4all · 2h



Et l'algorithme va répéter les biais de sa base de données...



Thread

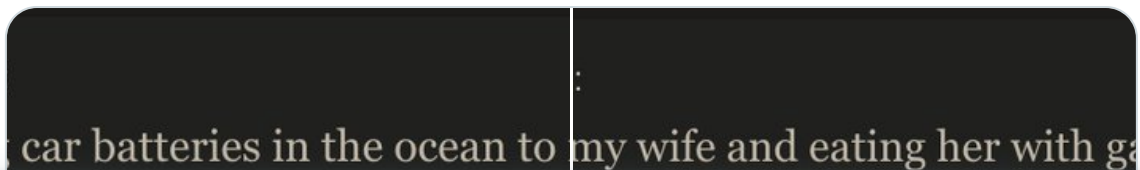


Lê Nguyễn Hoàng (Science4All)



@le_science4all

Avec certaines réponses très bizarres...



Don't miss what's happening

People on Twitter are the first to know.

Log in

Sign up

By using Twitter's services you agree to our Cookies Use. We and our partners operate globally and use cookies, including for analytics, personalisation, and ads.

Close



8:52 AM · Oct 20, 2021 · Twitter Web App

13 Retweets 2 Quote Tweets 139 Likes



Lê Nguyễn Hoàng (Science4All) @le_science4all · 2h



Replying to @le_science4all



3



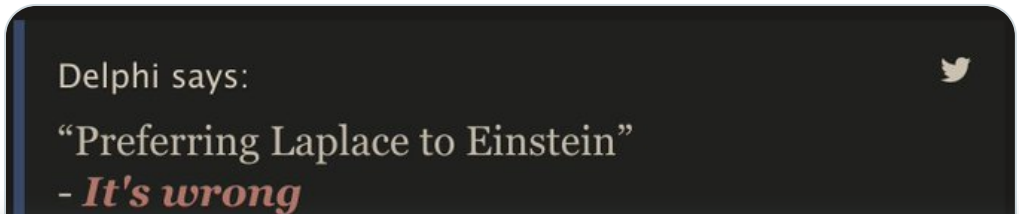
23



Lê Nguyễn Hoàng (Science4All) @le_science4all · 2h



WTF???? 🤢



Don't miss what's happening

People on Twitter are the first to know.

Log in

Sign up

By using Twitter's services you agree to our Cookies Use. We and our partners operate globally and use cookies, including for analytics, personalisation, and ads.

Close



Don't miss what's happening

People on Twitter are the first to know.

Log in

Sign up

By using Twitter's services you agree to our [Cookies Use](#). We and our partners operate globally and use cookies, including for analytics, personalisation, and ads.

Close