



IN NEED OF GUARDRAILS

OpenAI 最高経営責任者

サム・アルトマン

AIの安全性と有用性

対話型生成 AI の ChatGPT を生んだ企業 OpenAI の最高経営責任者サム・アルトマン。人類全体に利益をもたらす汎用人工知能 (AGI) を普及・発展させることを目標に、AI 分野の研究を最前線で行う彼が、今年1月の世界経済フォーラム年次総会 (通称「ダボス会議」) で、AI の安全性と有用性について語った。



インタビュアー

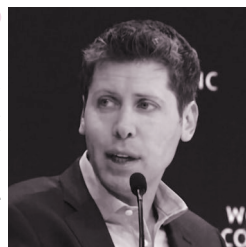
ファリード・ザカリア

インド出身のジャーナリスト、国際問題評論家。イェール大学卒業後、ハーバード大学で博士号を取得。国際政治経済ジャーナル「フォーリン・アフェアーズ」編集長、ニュース週刊誌「ニュースウィーク」の国際版編集長を経て、2008年6月よりCNNで「Fareed Zakaria GPS」の番組ホストを務める。1964年、ムンバイ生まれ。

ゲスト

サム・アルトマン

アメリカの起業家、投資家、プログラマー。OpenAI社の最高経営責任者。ベンチャーキャピタルのYコンビネータ社の元代表。スタンフォード大学でコンピューターサイエンスを学び、2005年に退学。2017年に、数々のスタートアップ企業を支援してきた功績をたたえられ、カナダのウオータールー大学から名誉学位を授与される。1985年生まれ。



Listening Quiz

82 AIのブラックボックス問題

Fareed Zakaria The thing that I think people worry about is the ability to trust AI. And part of that trust, I think, always comes when you understand how it works. And one of the problems is figuring out why it does what it does. Do you think that we will get there, or is it getting so inherently complicated that we are at some level just going to have to trust the black box?

Sam Altman I think humans are pretty forgiving of other humans making mistakes but not really at all forgiving of computers making mistakes. And so people who say things like, “Self-driving cars are already safer than human-driven cars”—it probably has to be safer by a factor of between 10 and 100 before people will accept it, maybe even more. And I think the same thing is going to happen for other AI systems.

I also think that what it means to verify or understand what’s going on is going to be a little bit different than people think right now. I actually can’t look in your brain and look at the 100 trillion synapses and try to understand what’s happening at each one and say, “OK, I really understand why he’s thinking what he’s thinking.” But what I can ask you to do is explain to me your reasoning. And you can explain, “First this, then this; then there’s this conclusion, then that one, and then there’s this.” And I can decide if that sounds reasonable to me or not. And I think our AI systems will also be able to do the same thing.

ability to do:

▶ここでは「～しても安心できること」という意味で使われている。

work:

①機能する ②《work with》～とともに働く、協働する

figure out:

①《wh-節とともに》～か解明する、理解する ②～を考え出す

inherently:

本質的に

complicated:

複雑な

at some level:

ある程度は

be forgiving of:

～に対して寛大である、寛容である

pretty:

かなり

self-driving car:

自動運転車

human-driven car:

人が運転する車

by a factor of:

～倍で

verify:

～を検証する

be going on:

進行している、起こっている

▶what is going onで「現在の状況、事態」の意味。

trillion:

兆

synapse:

シナプス ▶神経細胞であるニューロン間の接合部。

reasoning:

論理的思考、論理の筋道、論法

conclusion:

結論

reasonable:

妥当な、筋の通った、納得できる