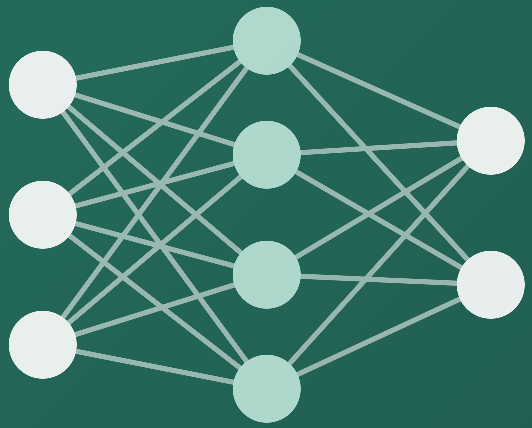
G検定対策講座　第２回

AIをめぐる議論

半世紀以上つづく、哲学者と科学者の論争



今日学ぶこと

1. **AIの誕生と判定**
   ダートマス会議・チューリングテスト
2. **強いAI と 弱いAI**
   サールが提唱した哲学的な議論
3. **AIの5つの難問**
   フレーム問題・シンボルグラウンディング ほか
4. **機械翻訳の進化**
   ルールベース → 統計的 → ニューラル
5. **シンギュラリティ**
   AIが人間の知能を超える瞬間

AIは本当に
「考えて」いる？

ChatGPTと話していると人間のよう。でも、理解しているのだろうか

1956年

ダートマス会議

人工知能という言葉が、初めて使われた研究会

主催は
ジョン・マッカーシー

具体例 いつ・どこで・何が

1956年、アメリカのダートマス大学で開かれた研究会

1956年
この年に開かれた

ダートマス大学
アメリカでの研究会

人工知能
この言葉が初めて使われた

主催したのは、ジョン・マッカーシーという研究者

ここが出る

AIが学問分野として誕生した

試験は**命名者・場・年**の3点セットで問われる

具体例 **3点セットで覚える**

試験は、この3つをセットで覚える

命名者	場	年
ジョン・マッカーシー	ダートマス会議	1956年

ミンスキー・シャノン・ニューウェル・サイモンらが参加

ここで AI が独立した学問分野として誕生した

1950年

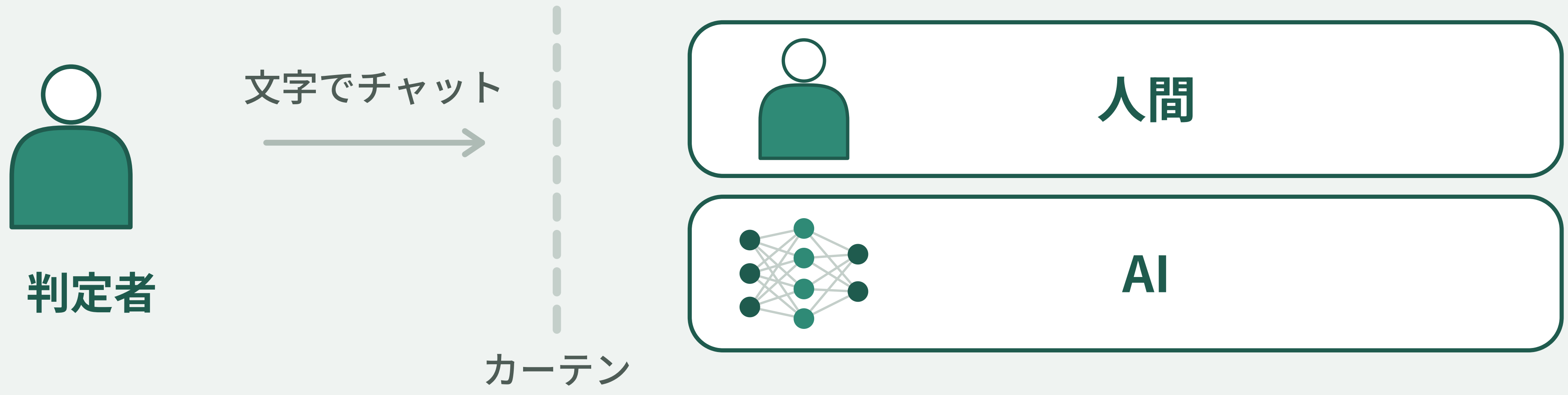
チューリングテスト

機械が**思考できるか**を判定する基準。アラン・チューリングが提案

具体例

判定のしかた

判定者が、カーテン越しに文字でチャットする



判定の考え方

見分けられなければ、思考しているとみなす

思考の中身には踏み込まず、**会話という外面だけ**で判定する

いまも AI 研究の
哲学的な基礎になっている

具体例

どこが画期的だったか

判定のしかた

見分けられた	見分けられない
相手が AI だと分かった 思考しているとはみなさない	人間か AI か分からない 思考しているとみなす

思考の中身には踏み込まず、会話という外面だけで判定する点が画期的だった

1991年

ローブナーコンテスト

チューリングテストを実際に行う大会。毎年おこなわれている

ヒュー・ローブナーが立ち上げた

具体例

これまでの結果

チューリングテストを、実際に行う大会

AI が人間のふりをする
会話で挑む

→

判定者をだませるか
毎年おこなわれる

→

完全な合格者はゼロ
まだ現れていない

人間と区別がつかない AI は、まだ実現していないということ

弱い AI

特定のタスクに特化したもの

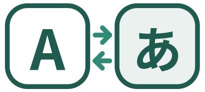
現在の AI はすべてこちら。分類はサールが1980年に提唱

具体例 **どんなものがあるか**

弱い AI —— 特定のタスクに特化したもの

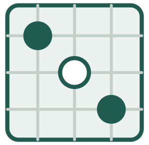
顔認証

顔を見分けるだけ



翻訳

訳すだけ



囲碁

打つだけ

ChatGPT・自動運転・AlphaGo。現在の AI はすべてこちら

強い AI（AGI）

何でもこなせる汎用の知能

人間のように、1つの AI ですべてをこなす

2026年現在
まだ実現していない

具体例 弱いAIとの違い

人間のように、何でもこなせる汎用の知能

- 考える
- 話す
- 作る

1つの AI が すべてをこなす

2026年現在、まだ実現していない

1980年

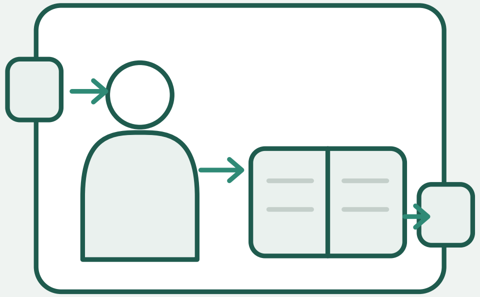
中国語の部屋

サールが示した、**強い AI への反論**となる思考実験

具体例

主張はシンプル

強い AI に反論するための、思考実験



ルールに従って記号を操作するだけでは「理解」とは言えない

答えを出せることと、理解していることは別だ

実験のシナリオ

質問を入れると、答えが返ってくる

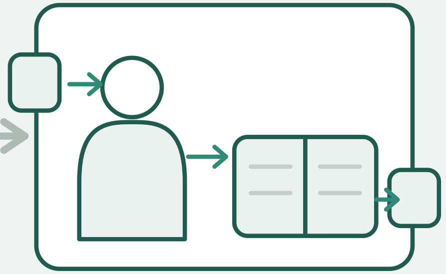
中の人は**中国語をまったく知らない**

具体例 部屋の中で起きていること

中の人は、中国語をまったく知らない

中国語の質問
外から入れる

→



→

中国語の答え
外へ返す

辞書とルールブックだけがある部屋

ルールブックに従って、答えを返しているだけ

この実験が示すこと

外から見れば、理解しているように見える

でも中の人は、**意味をまったく理解していない**

強い AI への
強力な反論になった

具体例 **外と中のズレ**

同じ部屋を、外と中から見ると

外から見れば	でも中の人は
中国語の答えが返ってくる	記号を照合しているだけ
理解しているように見える	意味をまったく理解していない

サールの主張は「機械も同じだ」。答えを出せる＝理解している、とは限らない

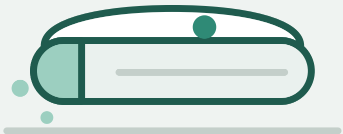
1969年

フレーム問題

マッカーシーとヘイズが提唱した**根本的な難問**

考慮すべきことが
無限にある

具体例 現実の世界で行動すると



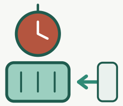
AI・ロボット

現実の世界で行動しようとすると

- 床は滑らないか
- 天井は崩れないか
- 車輪は回るか
- 壁の色は変わるか

… 無限に続く

何を考慮し、何を無視するかを決めるのが難しい



ロボットの寓話

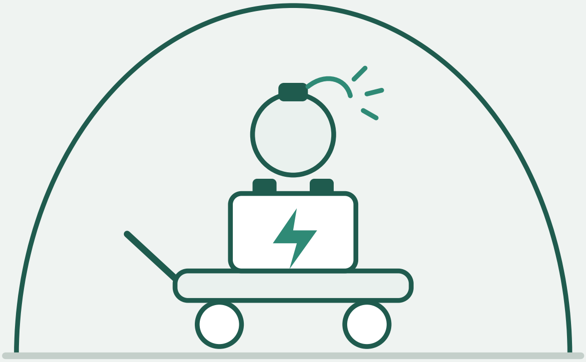
バッテリーの上に、時限爆弾

バッテリーだけを運び出すのが任務

具体例

設定

洞窟の中に、予備のバッテリーがある

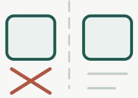


その上に、時限爆弾が乗っている

バッテリーだけを運び出すのが任務

ダニエル・デネットが1984年の論文で紹介した、有名な例

R1 と R1-D1



運び出したら爆発、全部考えたら動けない

副作用を考えないと失敗し、**全部考えると止まる**

具体例 **最初の2台**

最初の2台は、こう失敗した

R1	R1-D1
ワゴンごと運び出した 爆弾も一緒に運んで爆発	すべての副作用を考えた 多すぎて動けなくなった

副作用を考えないと失敗し、全部考えると止まってしまう

R2-D1 と結論

判定が終わらず、やはり動けない

知能とは**関係あることを見抜く力**

現代の AI でも
未解決の問題

具体例 **3台目と、この寓話の結論**

3台目も、やはり動けなかった

R2-D1

関係ないことを無視しようとしたが、判定が終わらない

↓

知能とは、何が関係あるかを直感で見抜く力

1990年

シンボルグラウンディング問題

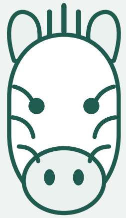
記号（言葉）と**現実の物事**をどう結びつけるか

具体例 **問いの中身**

記号（言葉）と、現実の物事をどう結びつけるか

「シマウマ」
という記号

どう結びつく？



現実のシマウマ

1990年、認知科学者スティーブン・ハーナッドが提唱した

シマウマの例

記号が記号を指す、堂々巡り

AI にはその経験がない。**中国語の部屋にとても近い議論**

具体例

辞書を引いてみると

辞書を引いても、また別の言葉が出てくるだけ

シマウマ → 白黒のしま模様の馬 → 白黒とは？ しまとは？

また別の言葉で説明される（堂々巡り）

人間は、見た・触った経験で言葉と物を結びつけている

第一次AIブーム

トイ・プロブレム

おもちゃのような問題しか解けない、と批判された

1950年代〜60年代に
AI が直面した限界

具体例 **何が解けて、何が解けなかったか**

歯が立たなかった問題と、解けた問題

✕ 歯が立たない	◯ 解けた
医療の診断・経営の判断 言葉の理解	迷路・ハノイの塔 簡単なパズル

条件が限られた問題しか解けず、これが第一次 AI 冬の時代を招いた

第二次AIブーム

知識獲得のボトルネック

専門家の知識を**入力しきれない**という限界

これが**機械学習**の
時代を呼び込んだ

具体例

どうやって知識を入れたか

専門家の知識を、ルールとして入力する方式

医師や弁護士
専門家の知識を

→

If-Then のルールに
もしこうならこうする

→

入力しきれない
知識が膨大すぎる

「歩くときは呼吸する」といった暗黙の常識までは書き出せない

1990年代〜

身体性

真の知能には**身体が必要**だという立場。ロドニー・ブルックスら

具体例 体感を通じて理解する

「重い」「熱い」を、どう理解するか

重い ／ 熱い
言葉としての概念

→

持ち上げる・触る

→

はじめて理解できる
記号の操作だけでは届かない

身体を持たない言語モデルは、本当に理解していると言えるのか

第1世代

ルールベース機械翻訳

文法の規則と辞書で訳す、古典的な方式。**1990年代まで**

具体例

機械翻訳の進化

機械翻訳は、3つの世代を経て進化してきた

第1世代　ルールベース機械翻訳
文法の規則と辞書で訳す　／　1990年代まで

第2世代

統計的機械翻訳

対訳データ（コーパス）から統計的に学ぶ。**IBM が先駆け**

具体例

機械翻訳の進化

機械翻訳は、3つの世代を経て進化してきた

第1世代　ルールベース機械翻訳
文法の規則と辞書で訳す　／　1990年代まで

第2世代　統計的機械翻訳
対訳データから統計的に学ぶ　／　IBM が先駆け

第3世代

ニューラル機械翻訳

ディープラーニングを使う。現在の Google 翻訳・DeepL

試験に出るのは
前の2つのほう

具体例

機械翻訳の進化

機械翻訳は、3つの世代を経て進化してきた

第1世代　ルールベース機械翻訳
文法の規則と辞書で訳す　／　1990年代まで

第2世代　統計的機械翻訳
対訳データから統計的に学ぶ　／　IBM が先駆け

第3世代　ニューラル機械翻訳
ディープラーニングを使う　／　Google 翻訳・DeepL

技術的特異点

シンギュラリティ

AI が人間の知能を超え、自己改良が爆発的に加速する瞬間

> 概念を広めたのは
> 数学者ヴィンジ（1993年）

具体例

何が起きる瞬間なのか

AI が人間の知能を超え、自己改良が爆発的に加速する

シンギュラリティ
技術的特異点

ここが出る

ヴィンジが概念、カーツワイルが2045年

楽観派は病気や貧困の解決を期待、悲観派は人類存続の危機を心配

具体例 2人の役割を区別する

2人の役割を、はっきり分けて覚える

1993年
ヴァーナー・ヴィンジ
数学者。概念を最初に広めた

2005年
レイ・カーツワイル
未来学者。2045年に到達と予測

この予測が、いわゆる2045年問題の由来

AIをめぐる議論のキーワード

試験頻出であると同時に、いまも続いている現在進行形の議論

- **誕生** **1956年 ダートマス会議（マッカーシーが命名）。判定はチューリングテスト**
- **強い / 弱い** サールの分類。現在の AI はすべて**弱い AI**。強い AI は未実現
- **中国語の部屋** 記号を操作するだけでは**理解とは言えない**という反論
- **5つの難問** **フレーム問題・シンボルグラウンディング・トイプロブレム・知識獲得のボトルネック・身体性**
- **機械翻訳** **ルールベース → 統計的 → ニューラル**の3世代
- **シンギュラリティ** **ヴィンジ**が概念、**カーツワイル**が2045年と予測

POINT 答えを出せることと、理解していること。この違いが議論の軸になっている