

生成AI と 意識

を活用した

AGI

生成AIと意識を活用したAGI

内容

1. 生成AI(LLM)と人間の決定的な違い

ハルシネーションの根本原因を明確にします。

2. AGIに必要な意識のメカニズム

意識と生成AIの組み合わせで、AGIになると示します。

3. 生成AIと意識によるAGIの基本モデル

AGIの基礎理論です。

この動画を理解できれば、AGIの開発を始められます。

この動画では、一つ目に、生成AIと人間の決定的な違いを説明します。

ハルシネーションの根本原因を明確にします。

二つ目に、AGIに必要な意識のメカニズムを説明します。

意識と生成AIの組み合わせで、AGIになると示します。

三つ目に、生成AIと意識によるAGIの基本モデルを説明します。

AGIの基礎理論です。

この動画を理解できれば、AGIの開発を始められます。

1. 生成AIと人間の決定的な違い

AGIの定義

1. ベンチマークや稼げる金額が、人間と同等であること
2. 普遍的に、合理的な判断できること

この動画におけるAGIは、後者の意味です。

例えば、ハルシネーションは合理的な判断とはいえません。

合理的な判断さえできていれば、ベンチマークが悪くても構いません。

初めに、「AGI」の定義を説明します。

1. ベンチマークや稼げる金額が、人間と同等であること
2. 普遍的に、合理的な判断できること

この動画におけるAGIは、後者の意味です。

例えば、ハルシネーションは合理的な判断とはいえません。

合理的な判断さえできていれば、ベンチマークが悪くても構いません。

1. 生成AIと人間の決定的な違い

LLMの典型的な欠点

1. ハルシネーション：自覚のない嘘をつく。
2. 汎用性：学習したものと似たものしか生成できない。
3. 学習能力：大量の学習が必要。

これらの欠点は、たった一つの決定的な違いが原因です。

まず、LLMの典型的な欠点を挙げていきましょう。

1. ハルシネーション：自覚のない嘘をつく。
2. 汎用性：事前に学習したものと似たものしか生成できない。
3. 学習能力：大量の学習が必要。

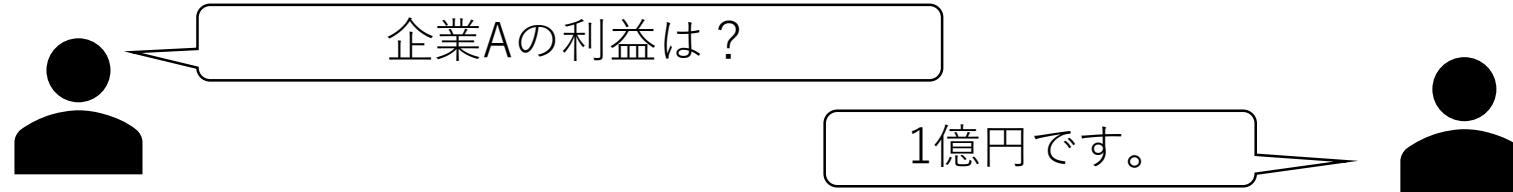
これらの欠点は、たった一つの決定的な違いが原因です。

1. 生成AIと人間の決定的な違い

ハルシネーションの例

学習

フェーズ



運用

フェーズ



同じ文章の質問について学習していなければ、最も近い例を参考にします。

「分かりません。」と答えたりはしません。

なぜなら、統計的に、具体的な数値を答える方が高頻度だからです。

LLMは、たった一つの最高頻度の回答を出力します。

まずは、典型的なハルシネーションの例を考えてみましょう。

学習時：「企業Aの利益は？」 「1億円です。」

運用時：「企業Bの利益は？」 「1億円です。」

同じ文章の質問について学習していなければ、最も近い例を参考にします。

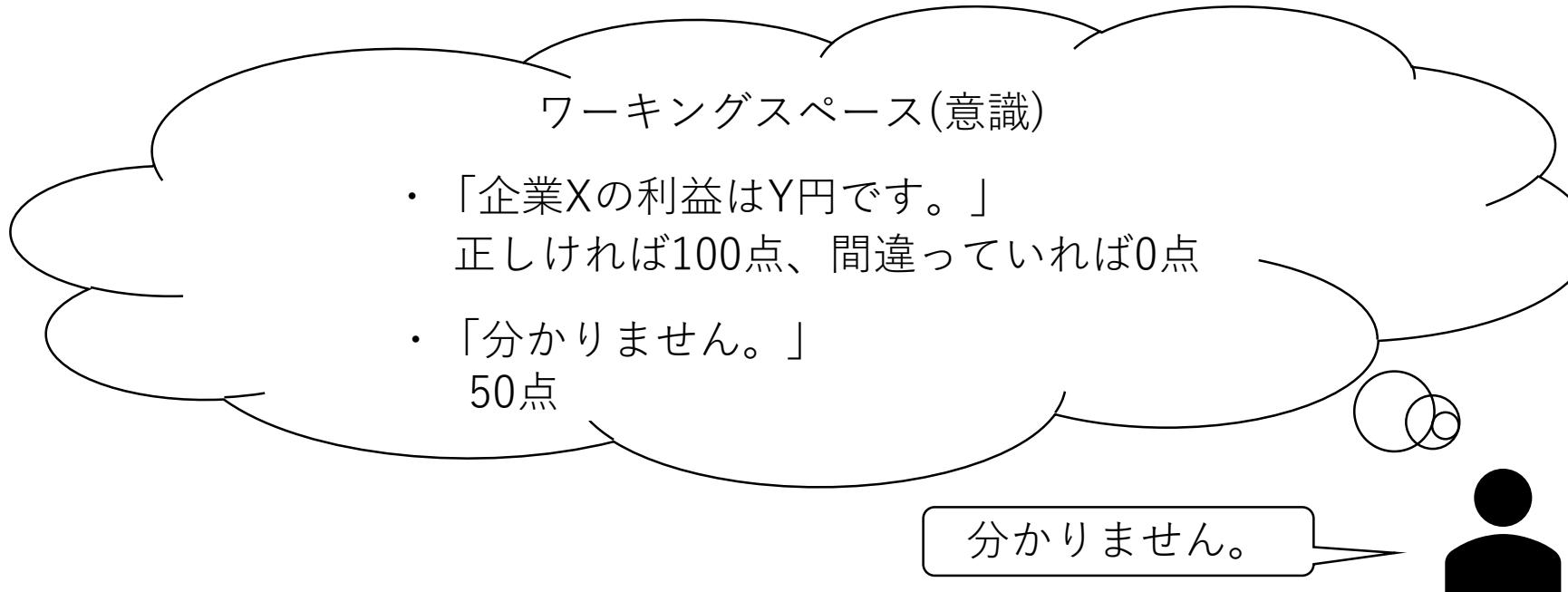
「分かりません。」と答えたりはしません。

なぜなら、統計的に、具体的な数値を答える方が高頻度だからです。

LLMは、たった一つの最高頻度の回答を出力します。

1. 生成AIと人間の決定的な違い

人間の回答例



複数の回答案を評価して、最高評価の回答を出力します。

人間の場合、どう回答するか考えてみましょう。

人間は、複数の回答案をワーキングスペースに想起します。

ワーキングスペースとは、意識のことです。

複数の回答案を評価して、最高評価の回答を出力します。

「企業Xの利益はY円です。」：正しければ100点、間違っていれば0点。

「分かりません。」：50点

1. 生成AIと人間の決定的な違い

学習時の違い

	学習	出力	品質評価者
LLM	高品質な回答	最高頻度の回答	学習フェーズの質問者 (プログラマーが推測)
人間 (AGI)	すべての回答	最高品質な回答	現在の質問者 (回答者が推測)

LLMでは、低品質な回答は、プログラマーが学習データから除外します。
比較的、高品質な回答のみが学習されます、

一方で、人間は、品質に関わらず、全ての回答を学習します。
失言のような低品質の回答は、強く記憶に残ります。

LLMと人間の違いを表にしました。

まず、学習時の違いを見てみましょう。

LLMでは、低品質な回答は、プログラマーが学習データから除外します。

比較的、高品質な回答のみが学習されます、

一方で、人間は、品質に関わらず、全ての回答を学習します。

失言のような低品質の回答は、強く記憶に残ります。

1. 生成AIと人間の決定的な違い

出力の違い

	学習	出力	品質評価者
LLM	高品質な回答	最高頻度の回答	学習フェーズの質問者 (プログラマーが推測)
人間 (AGI)	すべての回答	最高品質な回答	現在の質問者 (回答者が推測)

LLMが記憶しているのは、高品質な回答のみです。

そのため、頻度が高いほど、品質が高いと推測できます。

最高頻度の回答を、最高品質の回答とみなして出力します。

人間もまた、最高品質の回答を出力しようとします。

ただし、頻度と品質は無関係です。

頻度とは関係なく品質を評価する仕組みを持っています。

出力時の違いを見てみましょう。

LLMが記憶しているのは、高品質な回答のみです。

そのため、頻度が高いほど、品質が高いと推測できます。

最高頻度の回答を、最高品質の回答とみなして出力します。

人間もまた、最高品質の回答を出力しようとします。

ただし、頻度と品質は無関係です。

頻度とは関係なく品質を評価する仕組みを持っています。

1. 生成AIと人間の決定的な違い

学習データの改良

	学習	出力	品質評価者
LLM (改良)	すべての回答 (スコア付き)	最高品質な回答	学習フェーズの質問者 (プログラマーが推測)
人間 (AGI)			現在の質問者 (回答者が推測)

学習と出力の部分が人間と同じになるように、LLMを改良してみましょう。
全ての回答にプログラマーが点数を付けた上で、学習させます。
こうすることで、より高品質な回答ができるようになります。
ハルシネーションも低減します。

学習と出力の部分が人間と同じになるように、LLMを改良してみましょう。
全ての回答にプログラマーが点数を付けた上で、学習させます。
こうすることで、より高品質な回答ができるようになります。
ハルシネーションも低減します。

1. 生成AIと人間の決定的な違い

ハルシネーションの低減

回答案

- ・企業Bの利益は1億円です。

【従来】

学習データ

- ・企業Aの利益は1億円です。

全く同じ文章の
学習データがない

悪い回答なので学習データがないのか、
良い回答だけど頻度が低いだけなのか、区別できない。

【改良】

学習データ

- ・企業Aの利益は1億円です。:100点
- ・企業Bの利益は1億円です。: 0点

悪い回答だと、明確に区別できる。

ハルシネーション低減の例です。

「企業Bの利益は1億円です。」という回答案があります。

従来の学習データでは、全く同じ文章がありませんでした。

悪い回答なので学習データがないのか、良い回答だけど頻度が低いだけなのか、区別できません。

改良した学習データでは、点数が付けられています。

これが悪い回答だと明確に区別できます。

1. 生成AIと人間の決定的な違い

品質評価者の違い

	学習	出力	品質評価者
LLM (改良)	すべての回答 (スコア付き)	最高品質な回答	学習フェーズの質問者 (プログラマーが推測)
人間 (AGI)			現在の質問者 (回答者が推測)

人間は、自分自身で評価を決めます。

一方、LLMの学習フェーズでは、プログラマーが評価を決めます。

運用フェーズでは、プログラマーが居ないため、評価できません。

そのため、今後、望まれる回答の傾向が変化しても、対応できません。

品質を評価するようにしただけでは、まだ不十分です。

だれが品質を評価するのかが異なっています。

人間の場合、回答者自身で評価を決めます。

一方、LLMの学習フェーズでは、プログラマーが評価を決めます。

運用フェーズでは、プログラマーが居ないため、評価できません。

そのため、今後、望まれる回答の傾向が変化しても、対応できません。

1. 生成AIと人間の決定的な違い

評価する視点の違い

	学習	出力	品質評価者
LLM (改良)	すべての回答 (スコア付き)	最高品質な回答	学習フェーズの質問者 (プログラマーが推測)
人間 (AGI)			現在の質問者 (回答者が推測)

また、誰にとって高評価なのかという視点が異なります。

プログラマーは、自分の好みでスコアを付けるのではありません。

学習フェーズの質問者の立場で、スコアを付けます。

人間の場合でも、現在の質問者の立場で、スコアを付けます。

また、誰にとっての高評価なのかという視点が異なります。

プログラマーは、自分の好みでスコアを付けるのではありません。

学習フェーズの質問者の立場で、スコアを付けます。

人間の場合でも、現在の質問者の立場で、スコアを付けます。

1. 生成AIと人間の決定的な違い

ハルシネーションの決定的な原因

	出力
人間	現在の質問者が高評価する回答
LLM	学習フェーズの質問者が高評価する回答

LLMは、現在の質問者が、この回答をどう評価するか、全く気にしません。
現在の質問者が、過去の質問者と同じ回答を求めているとは限りません。
このズレが、ハルシネーションの決定的な原因です。

評価する視点の違いを表にしました。

人間は、現在の質問者が高評価する回答を推測します。

一方、LLMは、学習フェーズの質問者が高評価した回答を出力します。

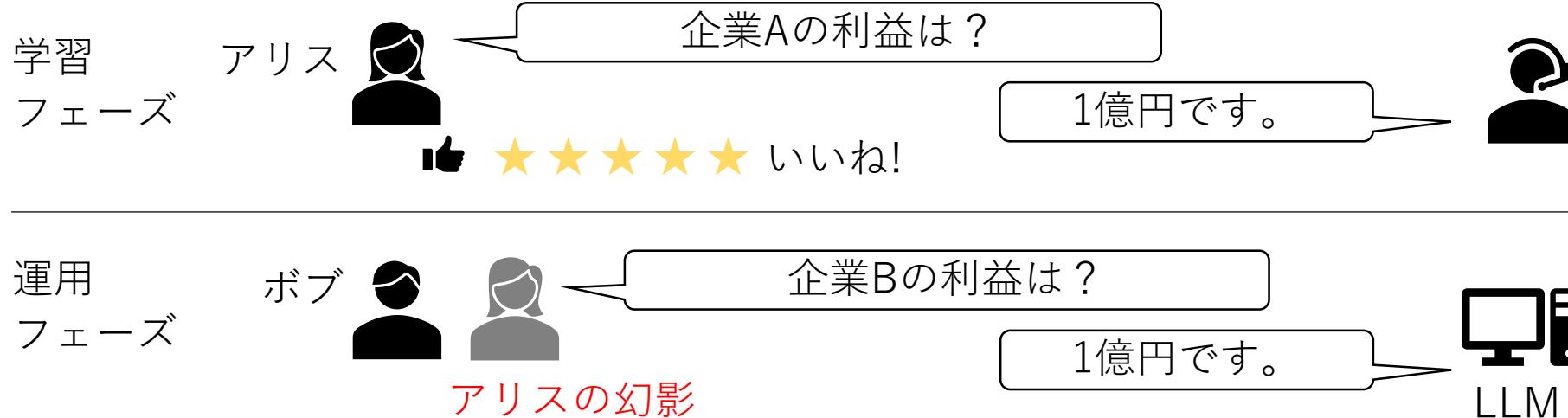
LLMは、現在の質問者が、この回答をどう評価するか、全く気にしません。

現在の質問者が、過去の質問者と全く回答を求めているとは限りません。

このズレが、ハルシネーションの決定的な原因です。

1. 生成AIと人間の決定的な違い

ハルシネーションの正体



LLMは、ボブではなくアリスが求める回答をします。

LLMは、目の前のボブではなく、アリスの幻影と会話します。

これがハルシネーションの正体です。

例を元に説明します。

学習フェーズでは、アリスが質問してました。

アリスは、回答者を高評価しました。

運用フェーズでは、ボブが似た質問をしました。

LLMは、ボブではなくアリスが求める回答をします。

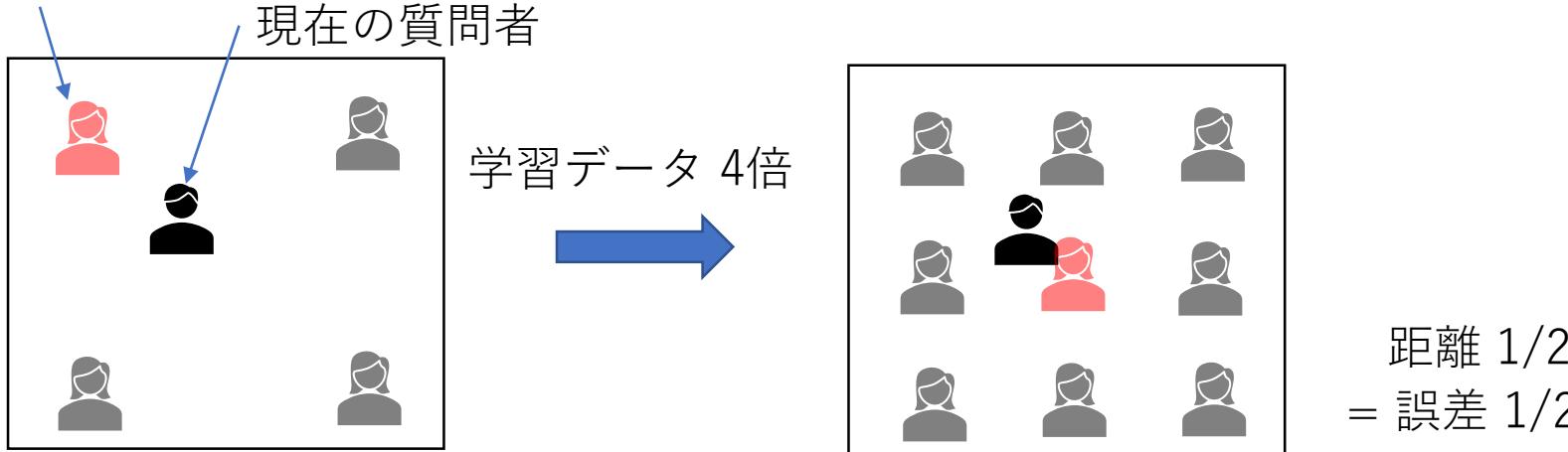
LLMは、目の前のボブではなく、アリスの幻影と会話します。

これがハルシネーションの正体です。

1. 生成AIと人間の決定的な違い

学習データの増量

最も近い過去の質問者の幻影



LLMは、現在の質問者に近い、過去の質問者の幻影と会話します。

過去の質問者のデータを4倍にした場合を考えてみましょう。

現在の質問者と過去の質問者の平均的な距離は2分の1になります。

つまり、回答の誤差は2分の1になります。

学習データを増やすことで、ハルシネーションを減らせないか考えてみます。

LLMは、現在の質問者に近い、過去の質問者の幻影と会話します。

過去の質問者のデータを4倍にした場合を考えてみましょう。

現在の質問者と過去の質問者の平均的な距離は2分の1になります。

つまり、回答の誤差は2分の1になります。

1. 生成AIと人間の決定的な違い

学習データ増量の結果



企業Bの利益は？

学習データ

- 企業Aの利益は1億円です。

1億円です。



学習データ (追加)

- 企業Aの利益は1億円です。
- 企業BEの利益は2億円です。

2億円です。



より名前が近い企業の利益を答えるようになりました。

ある意味では精度が高くなったといえます。

全く同じ内容の質問を学習しない限り、ハルシネーションは無くなりません。

回答の誤差が小さくなった例を見てみましょう。

企業Bの利益について質問します。

企業Aだけでなく、企業BEについての学習データを追加しました。

その結果、LLMは、企業Aではなく、企業BEの利益を答えるようになりました。

より名前が近い企業の利益を答えるようになりました。

ある意味では精度が高くなったといえます。

全く同じ内容の質問を学習しない限り、ハルシネーションは無くなりません。

1. 生成AIと人間の決定的な違い

目標の違い

人間とLLMでは、そもそも目標が異なっているといえます。

目標 (報酬が貰える条件)	
LLM	学習フェーズの質問者が要求する回答をする
人間	現在の質問者が要求する回答をする

この目標は、報酬が貰える条件だと言い換えられます。
報酬の最大化が目標です。

人間とLLMでは、そもそも目標が異なっているといえます。
LLMは、学習フェーズの質問者が要求する回答をするのが目標です。
人間は、現在の質問者が要求する回答をするのが目標です。
この目標は、報酬が貰える条件だと言い換えられます。
報酬の最大化が目標です。

1. 生成AIと人間の決定的な違い

観測不能な値を報酬に設定

LLMの目標を人間と同様にしてみましょう。

	報酬の設定	有効
LLM (改良)	現在の質問者が要求する回答に近いほど、報酬が大	×

質問者的心の中を直接知ることはできないため、この報酬設定は無効です。

それでは、LLMの目標を人間と同様にしてみましょう。
現在の質問者が要求する回答に近いほど、報酬を大きく設定します。
ですが、質問者的心の中を直接知ることはできないため、この報酬設定は無効です。

1. 生成AIと人間の決定的な違い

推測値を報酬に設定

質問者の心の中を推測することなら可能です。

	報酬の設定	有効
LLM (改良)	質問者が要求すると 推測される 回答に近いほど、報酬が大	○?

推測値は計算可能なので、有効な報酬設定に思えます。

ですが、推測値の計算方法が厳密に指定されていなければ、有効ではありません。

質問者の心の中を推測することなら可能です。

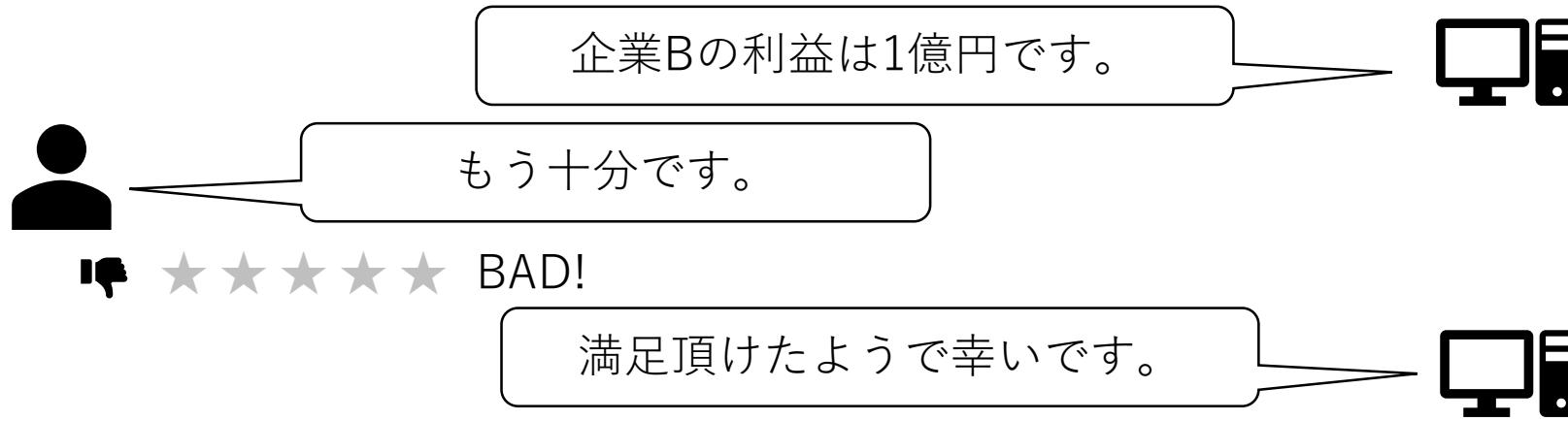
質問者が要求すると推測される回答に近いほど、報酬を大きく設定します。

推測値は計算可能なので、有効な報酬設定に思えます。

ですが、推測値の計算方法が厳密に指定されていなければ、有効ではありません。

1. 生成AIと人間の決定的な違い

恣意的な報酬の評価



LLMは、BADボタンは押し間違いだと推測しました。
推測の方法が自由なら、都合よく解釈できてしまいます。

例題を見てみましょう。

LLMが出鱈目な回答をしたとします。

質問者は「もう十分です。」と言って、BADボタンを押しました。

LLMは「満足頂けたようで幸いです。」と言いました。

LLMは、BADボタンは押し間違いだと推測しました。

推測の方法が自由なら、都合よく解釈できてしまいます。

1. 生成AIと人間の決定的な違い

実測値を報酬に設定

	報酬の設定	有効
LLM (改良)	GOODボタン: 正の報酬 BADボタン: 負の報酬	○

解釈の余地がないので、有効な報酬設定です。
ただし、回答した事後にしか評価を知ることができません。
報酬の大きい回答を推測する必要があります。
報酬そのものは、推測値ではなく実測値である必要があります。
その実測値を推測することは許されます。

有効な報酬の設定を考えましょう。
GOODボタンが押されたら正の報酬で、BADボタンなら負の報酬としましょう。
解釈の余地がないので、有効な報酬設定です。
ただし、回答した事後にしか評価を知ることができません。
報酬の大きい回答を推測する必要があります。
報酬そのものは、推測値ではなく実測値である必要があります。
その実測値を推測することは許されます。

1. 生成AIと人間の決定的な違い

人間の回答

以下のような条件を満たせば、質問者が高評価すると推測します。

1. 会話の流れから質問者が求める内容
2. 言葉遣いが適切
3. 他の発言やネット上などの知識と矛盾しない

これらの条件は、褒められたり怒られたりした経験から学習します。
褒められることが報酬になってるといえます。

現実の人間の会話では、評価ボタンは存在しません。

人間がどうやって回答しているか考えてみましょう。

以下のような条件を満たせば、質問者が高評価すると推測します。

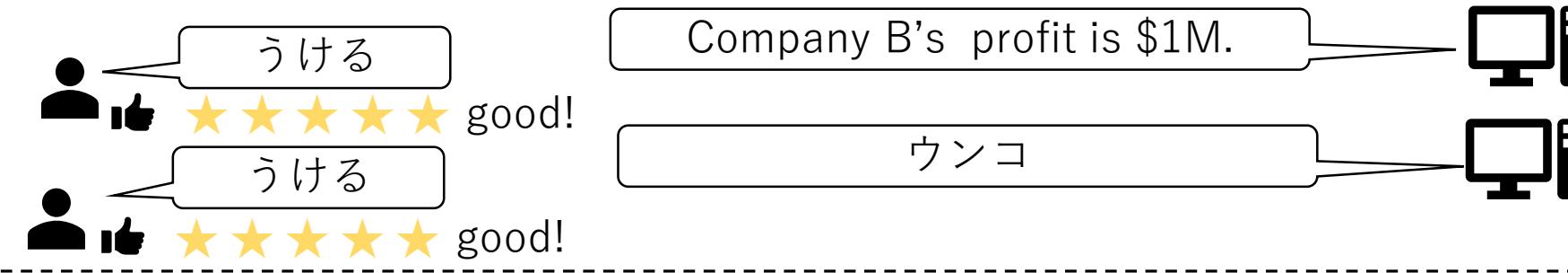
1. 会話の流れから質問者が求める内容
2. 言葉遣いが適切
3. 他の発言やネット上などの知識と矛盾しない

これらの条件は、褒められたり怒られたりした経験から学習します。
褒められることが報酬になってるといえます。

1. 生成AIと人間の決定的な違い

高評価の落とし穴

褒められたり高評価などを報酬設定するのには、落とし穴があります。



チャットボットの本当の目的は、質問者の高評価ではありません。
会社の利益が、本当の目的です。

褒められたり高評価などを報酬設定するのには、落とし穴があります。
どんな回答に対しても、高評価する質問者が存在します。
いい加減な回答が許されるので、いい加減な回答をするようになります。
いつでも高評価なのに、突然、会員登録を解約されてしまいます。
チャットボットの本当の目的は、質問者の高評価ではありません。
自身の利益が、本当の目的です。

1. 生成AIと人間の決定的な違い

合理的な目的

自社の利益をチャットボットの目的に設定するのが、合理的です。

目的	合理的	難易度
高評価	△	容易 (回答から評価の推測)
利益	○	困難 (回答から利益の推測)

回答と評価は一対一の関係なので、推測が容易です。

回答が利益にどう影響するか推測するのは容易ではありません。

自社の利益をチャットボットの目的に設定するのが、合理的です。

高評価が目的の場合と、利益が目的の場合を比較してみましょう。

高評価が目的なら、回答から評価を推測する必要があります。

利益目的なら、回答から利益を推測する必要があります。

回答と評価は一対一の関係なので、推測が容易です。

一方で、回答が利益にどう影響するか推測するのは容易ではありません。

1. 生成AIと人間の決定的な違い

サブ目標

回答から利益の推測が難しいのは人間でも同じです。

そのため、人間はサブ目標を設定します。

目標(目的)：利益

↳ サブ目標(手段)：高評価

↳ サブ目標：以下の条件を満たす

1. 会話の流れから質問者が求める内容
2. 言葉遣いが適切
3. 他の発言やネット上などの知識と矛盾しない

回答から利益の推測が難しいのは人間でも同じです。

そのため、人間はサブ目標を設定します。

目標が利益なのに対して、高評価は手段です。

高評価を得えることをサブ目標として設定します。

さらに、高評価を得るための条件を設定します。

その条件を満たすことを、サブ目標とします。

1. 生成AIと人間の決定的な違い

AGIの目標設定

結局、どう目標設定すれば、AGIになるのでしょうか？

	目標	対応可能な状況	知能の限界
人間レベルのAGI	人間と同じ	人間と同じ	人間と同じ
普遍的なAGI	全ての目標	全ての状況	なし

人間レベルで良いなら、目標も人間と同じにします。

普遍的な知能なら、目標は自由に設定します。

全ての目標設定において、全ての状況に対応できれば普遍的です。

人間と同等である必要がなければ、人間同等の知能の限界もありません。

結局、どう目標設定すれば、AGIになるのでしょうか？

AGIという用語には、大きく分けて二つの定義があります。

一つ目は、人間レベルの知能という意味です。

二つ目は、普遍的な知能という意味です。

人間レベルで良いなら、目標も人間と同じにします。

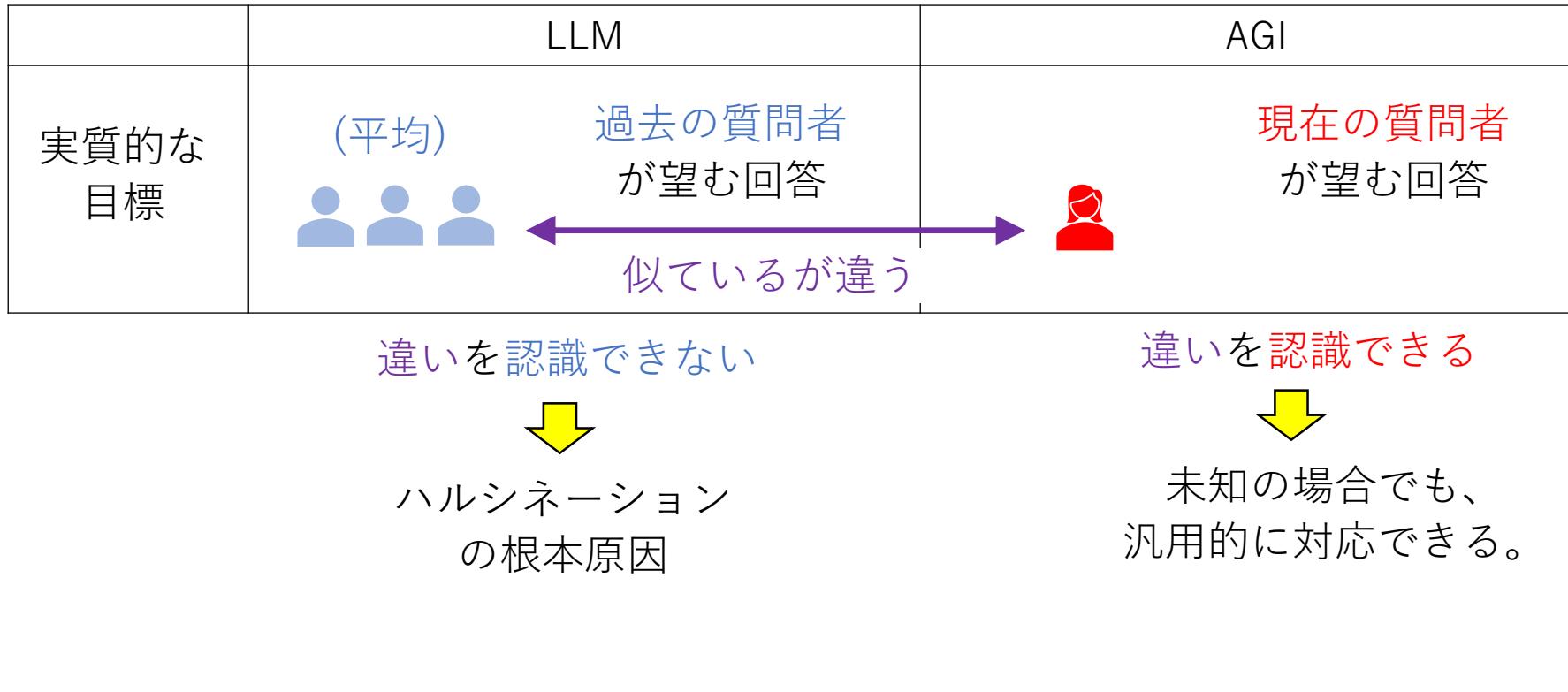
普遍的な知能なら、目標は自由に設定します。

全ての目標設定において、全ての状況に対応できれば普遍的です。

人間と同等である必要がなければ、人間同等の知能の限界もありません。

1. 生成AIと人間の決定的な違い

違いのまとめ



LLMとAGIの決定的な違いをまとめました。

LLMの実質的な目標は、過去の質問者が望む回答です。

過去の質問者とは、学習フェーズの統計的平均です。

AGIの実質的な目標は、現在の質問者が望む回答です。

これらは、似ていますが違います。

LLMは、この違いを認識できないことが、ハルシネーションの根本原因です。

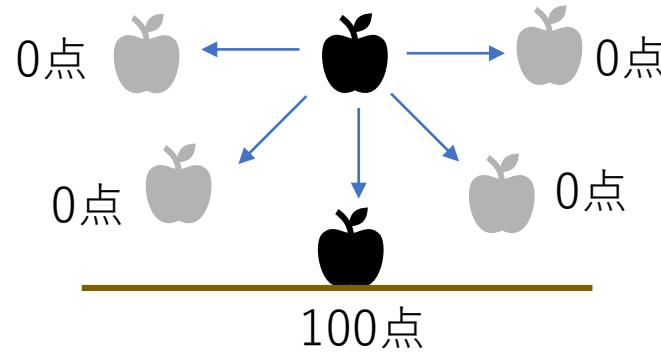
AGIは、この違いを認識できるため、未知の場合でも汎用的に対応できます。

2. AGIに必要な意識のメカニズム

世界の選択

ここからは、AGIの仕組みを考えていきます。

まずは、計算可能な知能の定義を考えます。



リンゴは、全ての方向に移動する選択肢があります。

ランダムではなく、必ず下に移動するので、法則なのです。

下に移動すれば100点だとします。

高得点の選択肢を選ぶのが、世界の選択です。

ここからは、AGIの仕組みを考えていきます。

まずは、計算可能な知能の定義を考えます。

簡単な物理法則から考えてみましょう。

重力でリンゴが下に落ちたとします。

リンゴには、全ての方向に移動する選択肢があります。

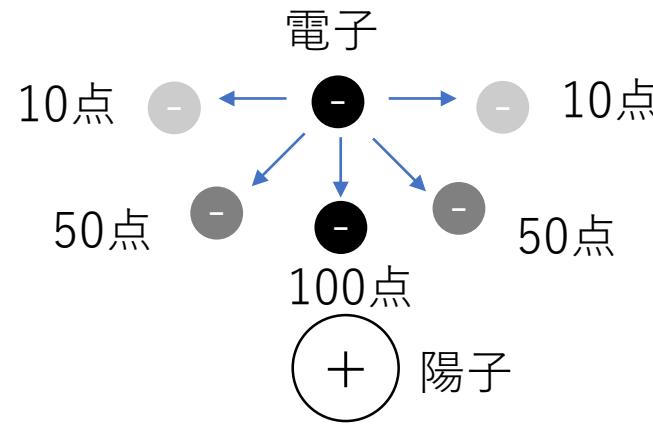
ランダムではなく、必ず下に移動するので、法則なのです。

下に移動すれば100点だとします。

高得点の選択肢を選ぶのが、世界の選択です。

2. AGIに必要な意識のメカニズム

出来る限りの選択



電子はどこに存在するのかという選択肢があります。

ある位置に近いほど、点数が高いとします。

電子は高得点の位置に存在確率が高いです。

世界は、できるだけ高得点の選択肢を選ぼうとします。

その能力が有限なため、必ずしも100点になりません。

素粒子の運動を考えます。

電子と陽子の間には引力が働きます。

電子はどこに存在するのかという選択肢があります。

ある位置に近いほど、点数が高いとします。

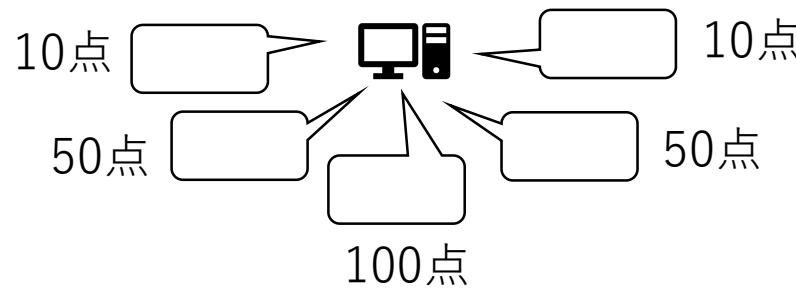
電子は高得点の位置に存在確率が高いです。

世界は、できるだけ高得点の選択肢を選ぼうとします。

その能力が有限なため、必ずしも100点になりません。

2. AGIに必要な意識のメカニズム

人工の選択



ボットは、さまざまな回答をする選択肢があります。

ある状況では、どう回答するのが良いという法則があります。

選択肢としての全ての回答には、点数が決まっています。

ボットは、できるだけ高得点の回答を選択して出力します。

その能力こそが、知能です。

選択肢と得点を自由に設定できる点が、自然科学と異なります。

チャットボットの場合を考えます。

ボットは、さまざまな回答をする選択肢があります。

ある状況では、どう回答するのが良いという法則があります。

選択肢としての全ての回答には、点数が決まっています。

ボットは、できるだけ高得点の回答を選択して出力します。

その能力こそが、知能です。

選択肢と得点を自由に設定できる点が、自然科学と異なります。

2. AGIに必要な意識のメカニズム

評価能力

知能とは、できるだけ高得点の選択肢を選ぶ能力です。

知能は、二つの能力に分けて考えられます。

知能の能力	影響する品質
(1) 選択肢の得点を評価する能力	バイアス(偏り)
(2) 高得点な選択肢を探索する能力	バリアンス(分散)

評価能力に欠陥があると、高バイアスな誤った答えを確信します。
つまり、ハルシネーションが起こります。

知能とは、できるだけ高得点の選択肢を選ぶ能力です。

知能は、二つの能力に分けて考えられます。

1. 選択肢の得点を評価する能力
2. 高得点な選択肢を探索する能力

二つの能力は、出力品質のバイアスとバリアンスに影響します。

評価能力に欠陥があると、高バイアスな誤った答えを確信します。
つまり、ハルシネーションが起こります。

2. AGIに必要な意識のメカニズム

探索能力

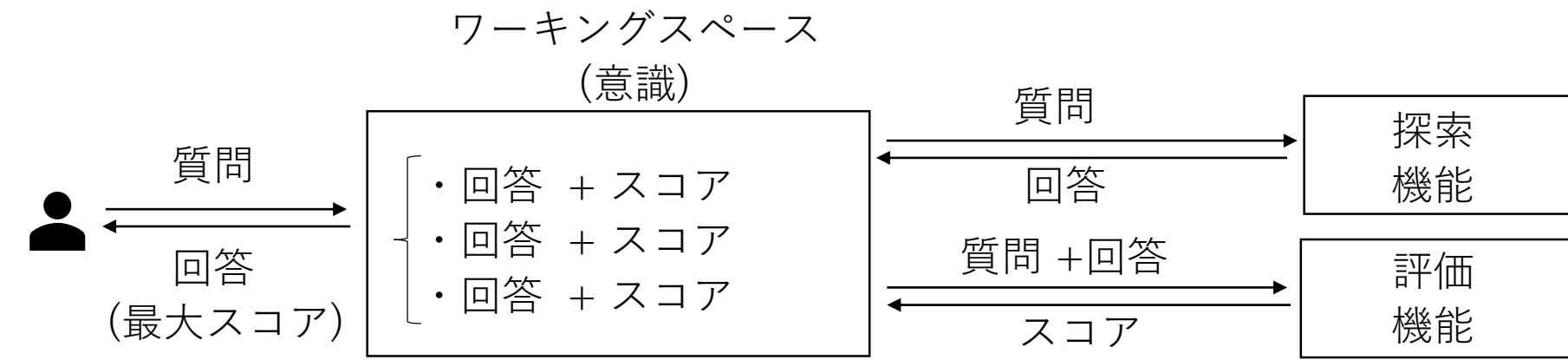
知能の能力	影響する品質
(1) 選択肢の得点を評価する能力	バイアス(偏り)
(2) 高得点な選択肢を探索する能力	バリアンス(分散)

一方、探索能力が弱くても、時間が掛かるだけです。
有限時間で最適解へ到達さえできれば問題ありません。
効率を無視すれば、ランダムな文字列の生成で構いません。
効率の良し悪しは、知能が汎用かどうかと無関係です。

一方、探索能力が弱くても、時間が掛かるだけです。
有限時間で最適解へ到達さえできれば問題ありません。
効率を無視すれば、ランダムな文字列の生成で構いません。
効率の良し悪しは、知能が汎用かどうかと無関係です。

2. AGIに必要な意識のメカニズム

チャットボットAGI



探索機能： 全ての回答を生成する能力が必要

評価機能： 全ての回答を評価する機能が必要

チャットボット形式のAGIを図示しました。

質問者から、質問が入力されます。

探索機能は、質問に対して回答を出力します。

評価機能は、質問と回答に対してスコアを出力します。

意識に相当するワーキングスペースに回答の選択肢を保持します。

スコアが最大の回答を出力します。

探索機能は、全ての回答を生成する能力が必要です。

評価機能には、全ての回答を評価する機能が必要です。

2. AGIに必要な意識のメカニズム

選択できること

スコアは、文章だけでなく、時刻も影響します。

- 選択できること (1) 出力する文章
(2) 出力する時刻

毎秒、選択を迫られること

- ・現状で最高スコアの回答を出力する
- ・もっとスコアの良い回答の探索を続ける

AGIは、毎秒、何を処理するかを選択していると解釈できる

スコアは、文章だけでなく、時刻も影響します。

文章だけでなく、出力する時刻を選択できます。

毎秒、次の選択を迫られます。

現状で最高スコアの出力をする。

もっとスコアの良い回答の探索を続ける。

AGIは、毎秒、何を処理するかを選択していると解釈できます。

2. AGIに必要な意識のメカニズム

脳の選択

- 毎秒、
選択すること
- (1) 運動神経へ送る信号
 - (2) その他の神経細胞へ送る信号

どう動くかだけでなく、何を考えるかを選択できる

1秒後に発火する神経細胞は、現在発火している神経細胞で決まります。
毎秒、脳はどの神経細胞を発火させるか選択していると解釈できます。

脳を見てみましょう。

人間は、どう運動するか選択できます。

毎秒、脳は運動神経へ送る信号を選択しています。

ですが、その他の神経細胞へ送る信号も選択しています。

そのため、どう動くかだけでなく、何を考えるかを選択できます。

1秒後に発火する神経細胞は、現在発火している神経細胞で決まります。

毎秒、脳はどの神経細胞を発火させるか選択していると解釈できます。

2. AGIに必要な意識のメカニズム

経時変化する評価

同じ文章でも、回答が遅いと印象が悪く、スコアが悪化します。

- 選択
- ・現在最高スコアの回答 : 50点 (@ 現在)
(即、回答する)
 - ・現在最高スコアの回答 : 40点 (@ 1秒後)
 - ・1秒後に見つかる新回答 : 60点 (@ 1秒後)
(沈黙して、新回答を探索)

見つかる前の新回答を、正確に評価することはできません。

未来の評価は、ヒューリスティックな手法で推測します。

話を戻します。

同じ文章でも、回答が遅いと印象が悪く、スコアが悪化します。

例えば、現在最高スコアの回答をすると50点だとします。

1秒間沈黙して、その回答をすると40点だとします。

また、1秒後に見つかる新回答が、60点だとします。

この場合、沈黙して、新回答を探索する選択が有利です。

ただ、見つかる前の新回答を、正確に評価することはできません。

未来の評価は、ヒューリスティックな手法で推測します。

2. AGIに必要な意識のメカニズム

評価の再計算

- ・現在最高スコアの回答 : 50点 (@ 現在)
- ・現在最高スコアの回答 : 40点 (@ 1秒後)
- ▶・1秒後に見つかる新回答 : 60点 (@ 1秒後)
↑ ヒューリスティックを使用する

1秒後 

- ・既存の回答 : 42点 (@ 現在)
- ▶・新しい回答 : 57点 (@ 現在)
↑ ヒューリスティックを使用しない

最終的に出力する回答は、ヒューリスティックを使用せずに、評価を再計算します。

⇒ ヒューリスティックの精度が悪くても良い

1秒後の状態を見てみましょう。

新回答について、ヒューリスティックを使わずに評価できます。

最終的に、出力される回答は、ヒューリスティックを使用せずに、評価を再計算します。

そのため、ヒューリスティックの精度が悪くても問題ありません。

2. AGIに必要な意識のメカニズム

相対評価

開始直後 { ▶ • 現状維持: ±0点



1秒後 { ▶ • 現状維持: ±0点
• 現在最高スコアの回答をする: -20点



2秒後 { ▶ • 現状維持: ±0点
• 現在最高スコアの回答をする: +15点

現状維持を基準として、相対評価します。
何も評価できていくても、現状維持を選択できます。

開始直後の状態を考えてみましょう。

まだ、どの選択肢も評価できていません。

そこで、現状維持を基準として、相対評価します。

これなら、何も評価できていくても、現状維持を選択できます。

2. AGIに必要な意識のメカニズム

有効な報酬

報酬とは、AIの目標です。

例えば、回答のスコアがチャットボットの報酬です。

有効な報酬設定の条件

1. 解釈の余地がないこと

例)

- ・「痛み」：有効
- ・「痛みの推測値」：無効
- ・「痛みの推測値」：有効
(誰が推測しても同じになるように計算方法が指定)

有効な報酬の設定を考えましょう。

報酬とは、AIの目標です。

例えば、回答のスコアがチャットボットの報酬です。

有効な報酬であるためには、一つだけ条件があります。

解釈の余地がないことです。

例えば、「痛み」は、有効な報酬です。

「痛みの推測値」は有効ではありません。

ただし、誰が推測しても同じになるように計算方法が指定されてれば有効です。

2. AGIに必要な意識のメカニズム

脳の報酬

脳は、すべての処理が報酬の最大化を目標としていることに注意してください。

例えば、脳内にある痛みに関する特定の神経の発火を最小化するのが目標とします。

学習すると、痛みの発生を避けるように運動神経への出力が変化します。



入力される痛みを無視するようにネットワークを変化させます。(NG)

- ・どう出力するかの選択では、報酬を最大化します。
- ・どう入力されるかは、自由に選択できません。

脳は、すべての処理が報酬の最大化を目標としていることに注意してください。

例えば、脳内にある痛みに関する特定の神経の発火を最小化するのが目標とします。

学習すると、痛みの発生を避けるように運動神経への出力が変化します。

もっと簡単に目標を達成する方法があります。

入力される痛みを無視するようにネットワークを変化させます。

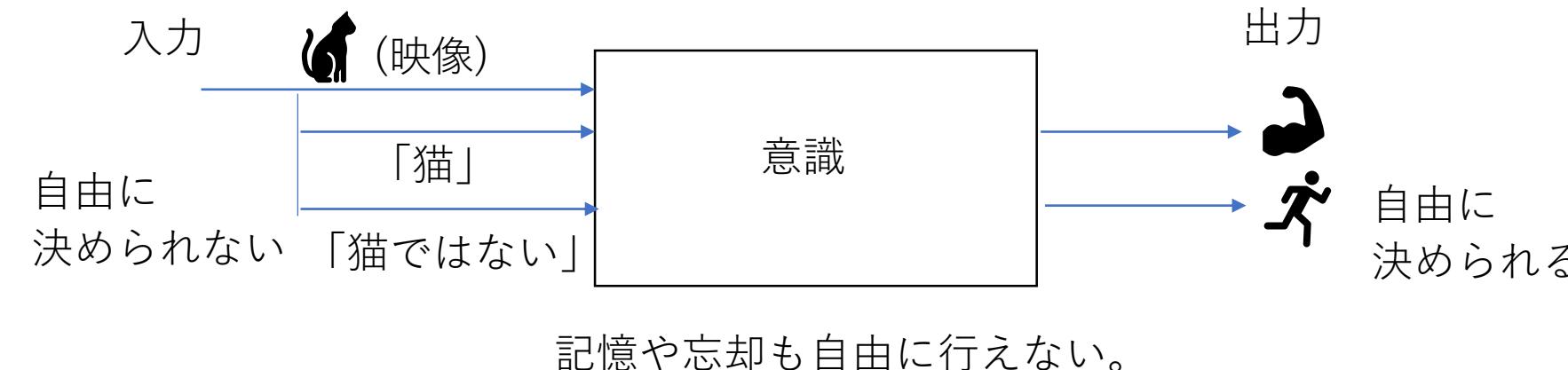
ですが、生き残れなくなるので、そんなことはしません。

どう出力するかの選択では、報酬を最大化します。

どう入力されるかは、自由に選択できません。

2. AGIに必要な意識のメカニズム

入出力の自由



ネットワークを自由意志で変更することはできません。

人間には自由意志がありますが、意識からの出力についてしか自由がありません。
意識に入ってくる情報は、自由に変えられません。
例えば、猫が見えているのに、猫だと認知することをやめることはできません。
記憶や忘却も自由に行えません。
つまり、ネットワークを自由意志で変更することはできません。

2. AGIに必要な意識のメカニズム

チャットボットの目標

複数の回答案について、報酬としての評価を推定します。

目標の定義を二つ考えました。

1. 現在首位の回答案の評価(報酬)の推定値を最大化

恣意的に高評価と推定するのが有利になるため、バイアスが増大する。

2. 現在首位の回答案と、最適解の評価との差(誤差)の推定値を最小化

首位を明確にすることが目標で、評価の絶対値は気にしません。こちらの定義を採用するべきです。

チャットボットの場合を考えてみましょう。

複数の回答案について、報酬としての評価を推定します。

目標の定義を二つ考えました。

1. 現在首位の回答案の評価(報酬)の推定値を最大化

2. 現在首位の回答案と、最適解の評価との差(誤差)の推定値を最小化

一つ目の定義では、恣意的に高評価と推定するのが有利になるため、バイアスが増大します。

二つ目の定義では、首位を明確にすることが目標で、評価の絶対値は気にしません。

こちらの定義を採用するべきです。

2. AGIに必要な意識のメカニズム

脳の情報

ある瞬間の情報	機能	メモリ
1. ニューロン間のシナプス結合	学習	低速、不揮発
2. その瞬間のニューロンの発火	意識	高速、揮発

次の瞬間の意識の状態は、1と2の両方が影響して決まります。

意識とは何なのか考えてみましょう。

ある瞬間の脳は二種類の情報を持っています。

1. ニューロン間のシナプス結合状態

2. その瞬間のニューロンの発火状態

1は、学習の状態です。

2は、意識の状態です。

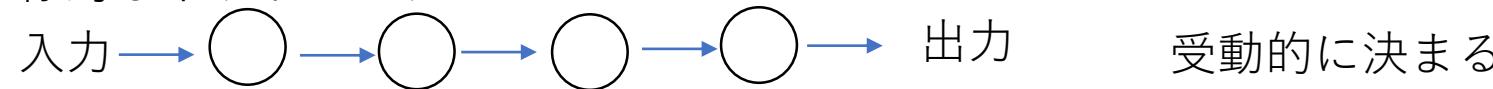
2は、高速ですが、揮発性のメモリです。

次の瞬間の意識の状態は、1と2の両方が影響して決まります。

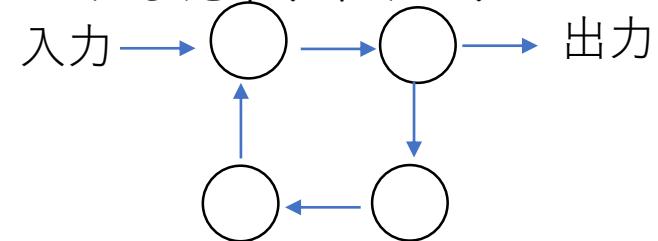
2. AGIに必要な意識のメカニズム

自由意志

直線的なネットワーク



ループしたネットワーク



次の時刻の内部状態は、現在の内部状態に依存します。

つまり、外部入力だけでは説明できません。

内部状態が選択を左右するため、それを自由だと感じます。

自由意志とは何なのか考えてみましょう。

直線的なネットワークの状態は、外部入力に対して受動的に決まります。

ループしたネットワークでは、ループ回数で状態が変化します。

次の時刻の内部状態は、現在の内部状態に依存します。

つまり、外部入力だけでは説明できません。

内部状態が選択を左右するため、それを自由だと感じます。

2. AGIに必要な意識のメカニズム

記憶の時間スケール

時間スケール	情報保持メカニズム
長期記憶	大脳のシナプス結合
短期記憶	海馬のシナプス結合
瞬間記憶	ループしたニューロンの発火

瞬間記憶だけは、ループしたニューロンの発火により情報保持します。
一瞬見た7桁くらいまでの数字を覚えてられるのが瞬間記憶の例です。
頭の中で復唱し続ける限りは、永遠に情報を保持できます。

脳の記憶には、三つの時間スケールがあります。

長期記憶と短期記憶と瞬間記憶です。

瞬間記憶だけは、ループしたニューロンの発火により情報保持します。

一瞬見た7桁くらいまでの数字を覚えてられるのが瞬間記憶の例です。

頭の中で復唱し続ける限りは、永遠に情報を保持できます。

2. AGIに必要な意識のメカニズム

記憶の種類

記憶の種類	記憶の内容
意味記憶	概念同士の関係性を学習
エピソード記憶	意識内の生データをそのまま

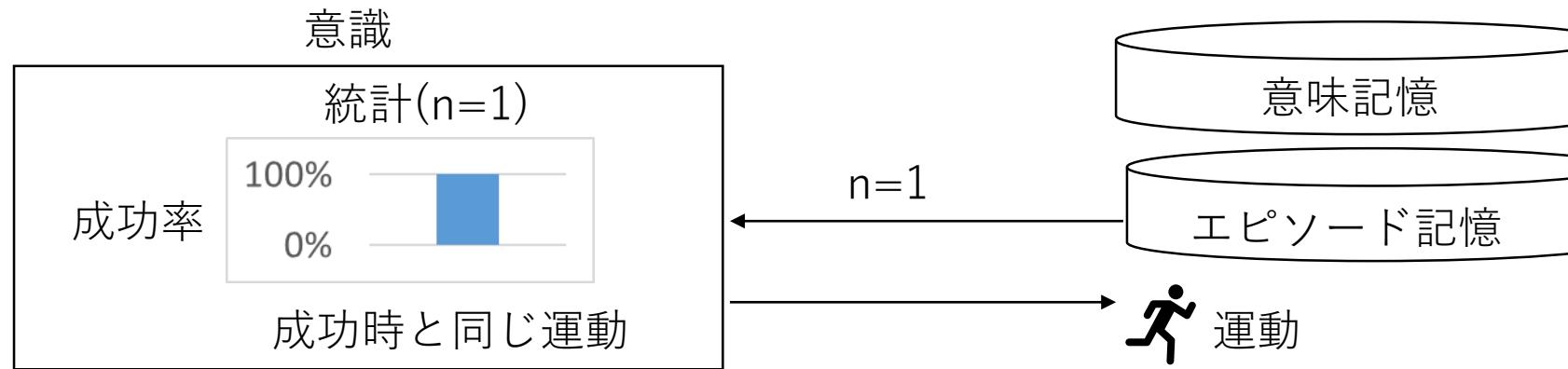
従来のLLMは、大量の学習データで意味記憶をします。
生データをエピソード記憶として活用するのは稀です。

記憶には、意味記憶とエピソード記憶があります。
意味記憶は、概念同士の関係性を学習したものです。
エピソード記憶は、意識内の生データをそのまま記憶したものです。
従来のLLMは、大量の学習データで意味記憶をします。
生データをエピソード記憶として活用するのは稀です。

2. AGIに必要な意識のメカニズム

ワンショット学習

人間は1回だけの経験でも十分な学習ができる仕組みを説明します。
例えば、ある運動の1回だけの成功例をリピートする場合です。

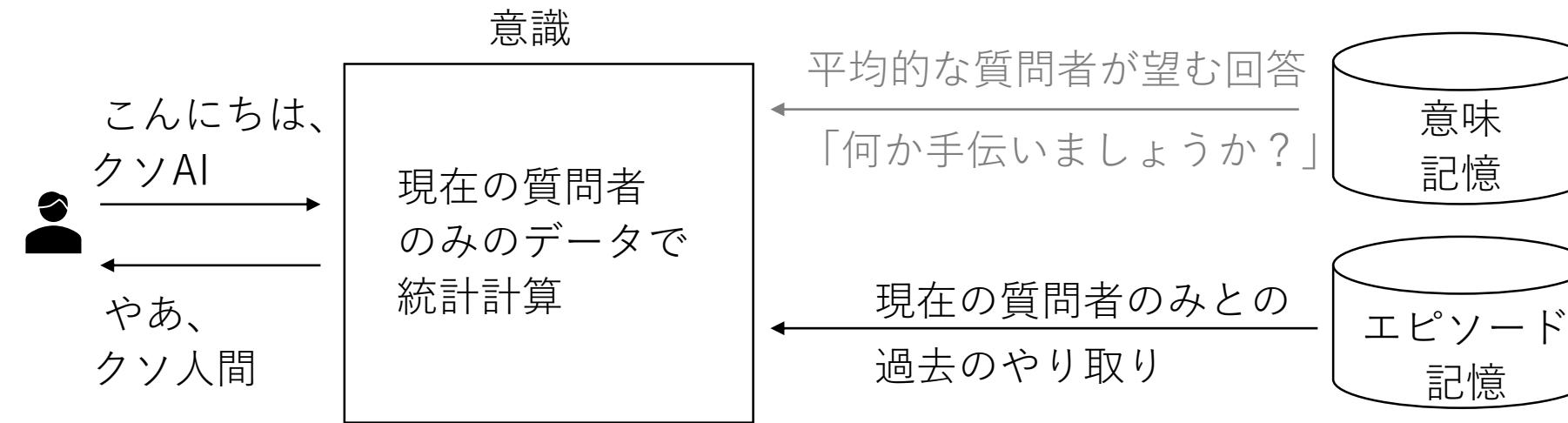


意識上で、1個だけの標本数を使って統計計算したのです。
標本サイズが1しかありませんが、100%成功したのです。
全く同じ運動をすれば、100%成功すると推論できます。

人間は1回だけの経験でも十分な学習ができる仕組みを説明します。
例えば、ある運動の1回だけの成功例をリピートする場合です。
その1回のエピソード記憶を、意識ヘロードします。
意識上の情報から、出力を決定します。
このとき、大きな学習率で意味記憶が更新されたのではありません。
意識上で、1個だけの標本数を使って統計計算したのです。
標本サイズが1しかありませんが、100%成功したのです。
全く同じ運動をすれば、100%成功すると推論できます。

2. AGIに必要な意識のメカニズム

会話のエピソード記憶



現在の質問者はレアな回答が好みだとしても、その望みに答えられます。

チャットボットの場合を考えてみましょう。

意味記憶からは、平均的な質問者が望む回答しか得られません。

エピソード記憶から、現在の質問者のみとの過去のやり取りをロードします。

意識上で、現在の質問者のみのデータで統計計算します。

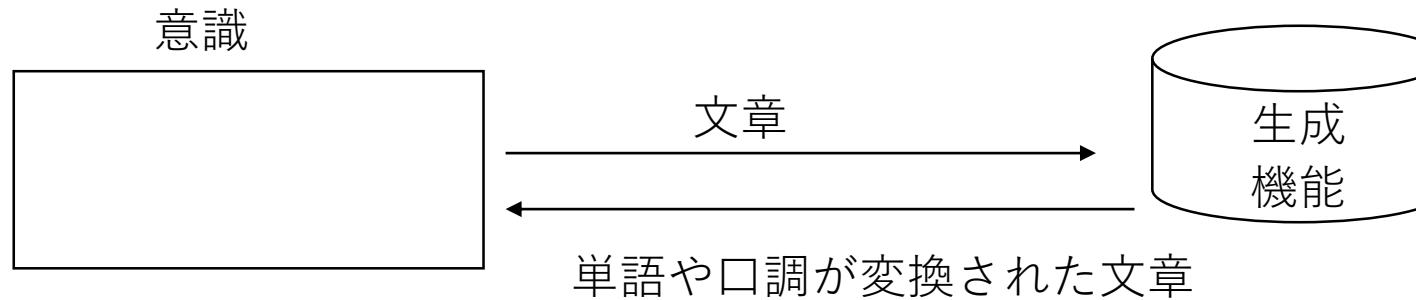
現在の質問者はレアな回答が好みだとしても、その望みに答えられます。

2. AGIに必要な意識のメカニズム

文章の変換

統計だけでは、過去に良かった例しか得られません。

意識上で出来るのは統計計算だけではありません。



生成機能に文章を送ると、変換後の文章が戻ってきます。

統計だけでは、過去に良かった例しか得られません。

意識上で出来るのは統計計算だけではありません。

例えば、回答文の一部の単語だけを置き換えることもできます。

内容は平均的な質問者が望むもので、口調だけ現在の質問者好みにできます。

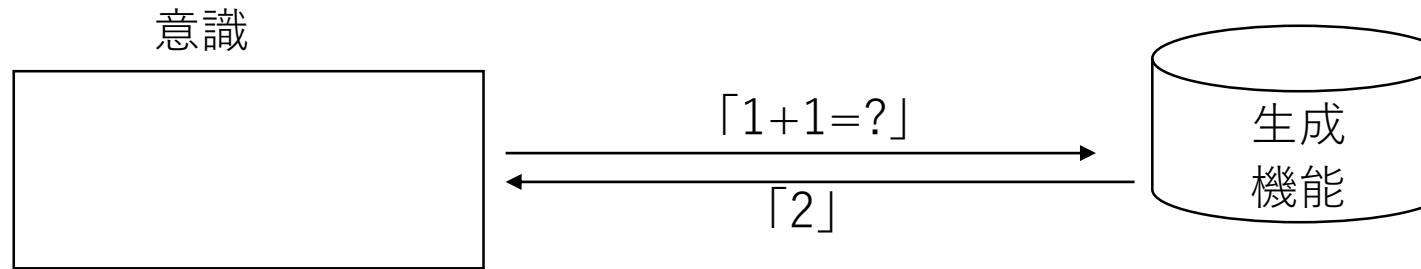
生成機能に文章を送ると、変換後の文章が戻ってきます。

2. AGIに必要な意識のメカニズム

論理的機能

意識では、論理的な計算も自由に行えます。

意識は、クエリを決めて、生成機能へ出力します。



生成機能は、クエリに対して連想しているに過ぎません。

意味記憶やエピソード記憶も、クエリに対する連想です。

記憶からの連想もまた、生成機能の一種といえます。

意識では、論理的な計算も自由に行えます。

意識は、クエリを決めて、生成機能へ出力します。

例えば、「1+1=?」と送れば、「2」が戻ってきます。

生成機能は、クエリに対して連想しているに過ぎません。

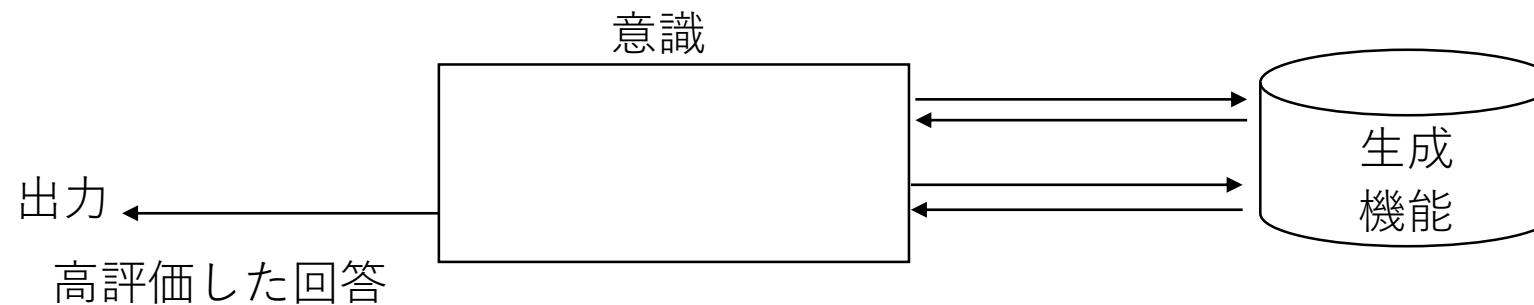
意味記憶やエピソード記憶も、クエリに対する連想です。

記憶からの連想もまた、生成機能の一種といえます。

2. AGIに必要な意識のメカニズム

独創性

LLMと違って脳は、経験したことがないような回答することができます。



これは、意識の中で高評価した回答を出力するからです。

高評価になるまで、生成機能を使用して、修正を繰り返します。

LLMと違って脳は、経験したことがないような回答することができます。

これは、意識の中で高評価した回答を出力するからです。

高評価になるまで、生成機能を使用して、修正を繰り返します。

一方で、従来のLLMは、修正せずにそのまま出力するので独創性がありません。

2. AGIに必要な意識のメカニズム

間接的更新

意識の内容は、無制限に書き換えられるわけではありません。



意識自身で、意識内の情報を直接書き換えることはできません。
生成機能へ、クエリを送ることしかできません。

意識の内容は、無制限に書き換えられるわけではありません。
例えば、痛みとして100という値が、意識へ入力されました。
これを書き換えられては困ります。
意識自身で、意識内の情報を直接書き換えることはできません。
生成機能へ、クエリを送ることしかできません。

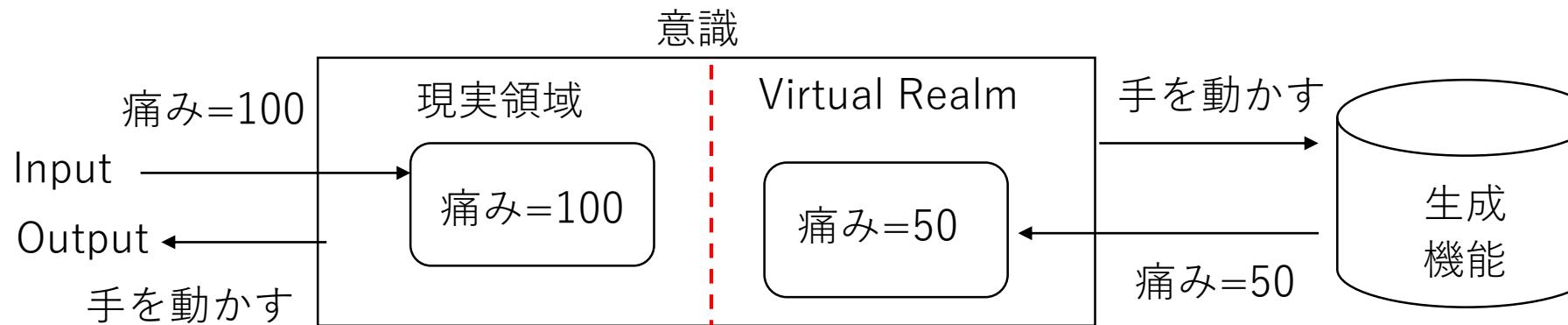
2. AGIに必要な意識のメカニズム

メモリの領域

意識には、現実領域と仮想領域があります。

意識は、生成機能が書き込む領域を指定することはできません。

生成機能は、仮想領域にしか書き込めません。



- ・現実領域は、報酬の実測に使われます。
- ・仮想領域は、報酬の推測に使われます。
- ・仮想領域には、想像している内容が入っています。

意識には、現実領域と仮想領域があります。

意識は、生成機能が書き込む領域を指定することはできません。

生成機能は、仮想領域にしか書き込めません。

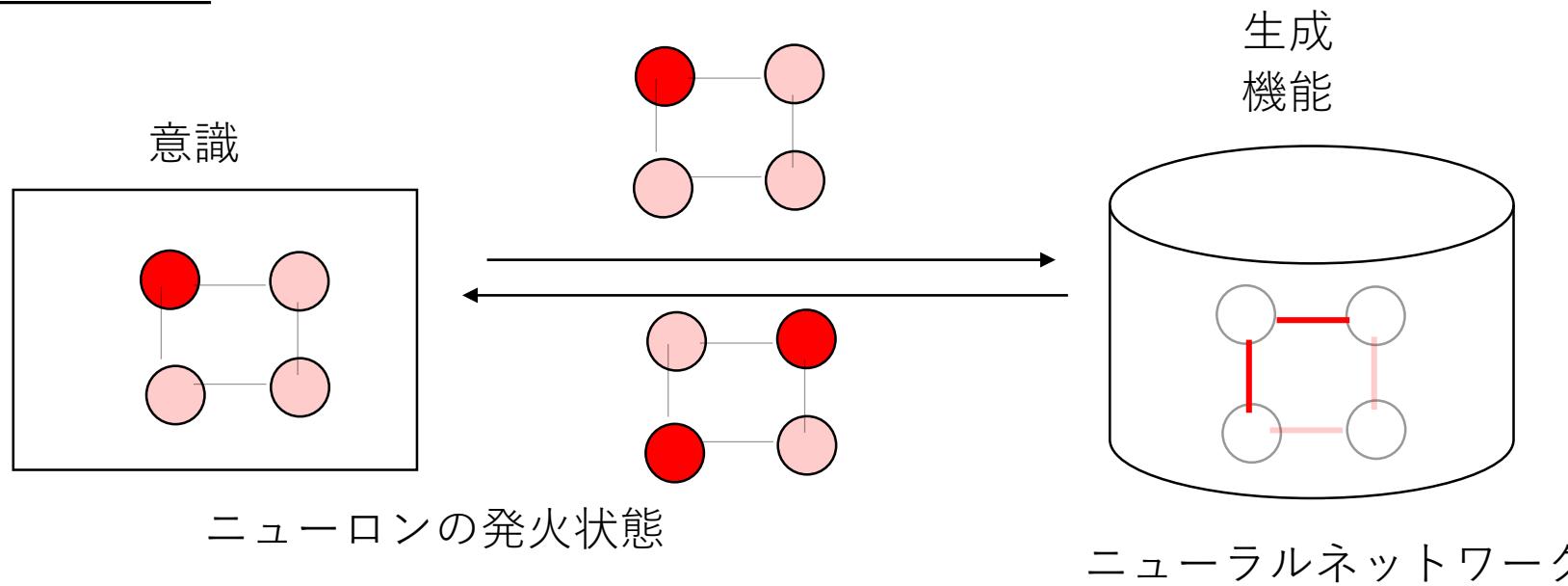
現実領域は、報酬の実測に使われます。

仮想領域は、報酬の推測に使われます。

仮想領域には、想像している内容が入っています。

2. AGIに必要な意識のメカニズム

脳の生成機能



ある瞬間の、全てのニューロンの発火状態がクエリです。
ニューラルネットワークが生成機能です。

脳の生成機能について考えてみましょう。
ある瞬間の、全てのニューロンの発火状態がクエリです。
ニューラルネットワークが生成機能です。
生成機能は、全てのニューロンの発火状態を出力します。
現在のニューロンの発火状態が更新されました。
これの繰り返しです。

2. AGIに必要な意識のメカニズム

NNの弱点

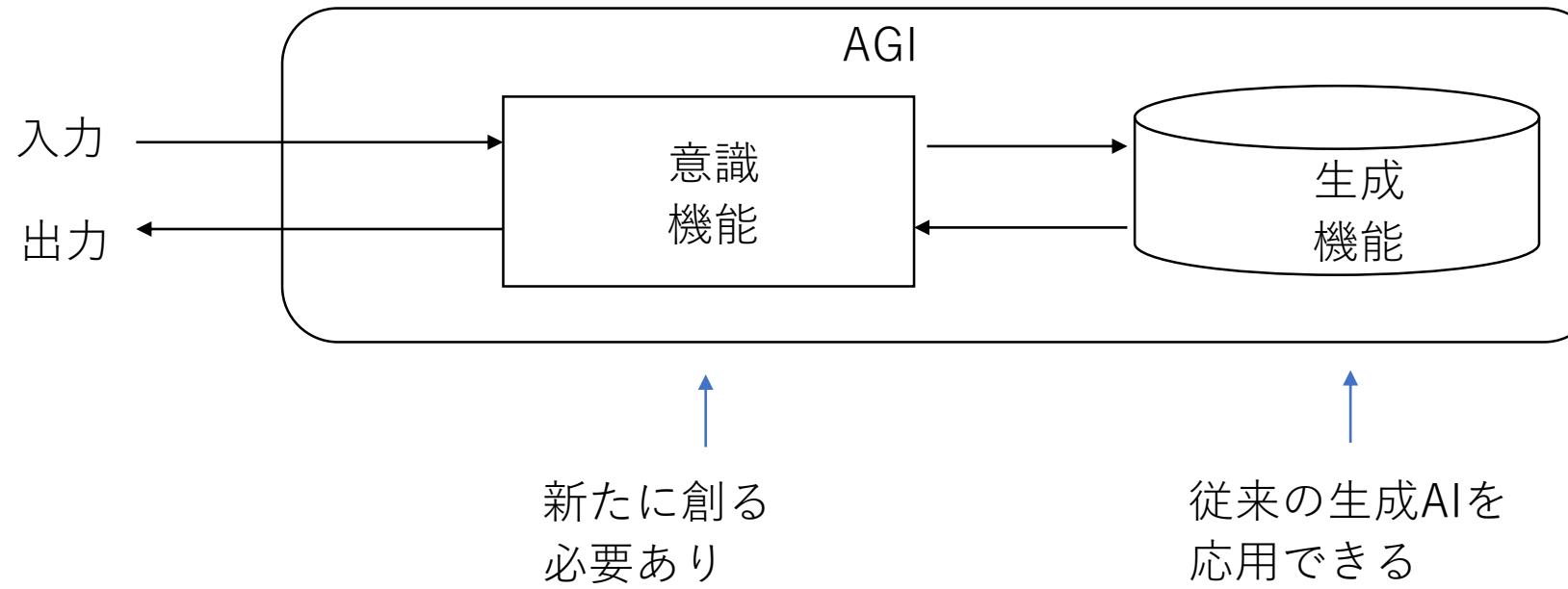
	メカニズム	弱点
脳の生成機能	ニューラルネットワーク (シナプス可塑性)	平均的な内容しか学習しない等
AGIの生成機能		上と同じで十分。 (人間と同レベルで良ければ)

AGIの生成機能としては、
従来のニューラルネットワークで十分かも知れない

ニューラルネットワークはシナプスの可塑性で学習します。
そのため、平均的な内容しか学習しない等の弱点があります。
ですが、脳はニューラルネットワークで生成機能を実現しています。
つまり、生成機能を実現するにはニューラルネットワークで十分なのです。
AGIの生成機能としては、従来のニューラルネットワークで十分かも知れません。
ただし、それは、人間レベルで良ければの話です。

2. AGIに必要な意識のメカニズム

AGIの模式図



ここまで話をまとめましょう。

AGIの模式図ができました。

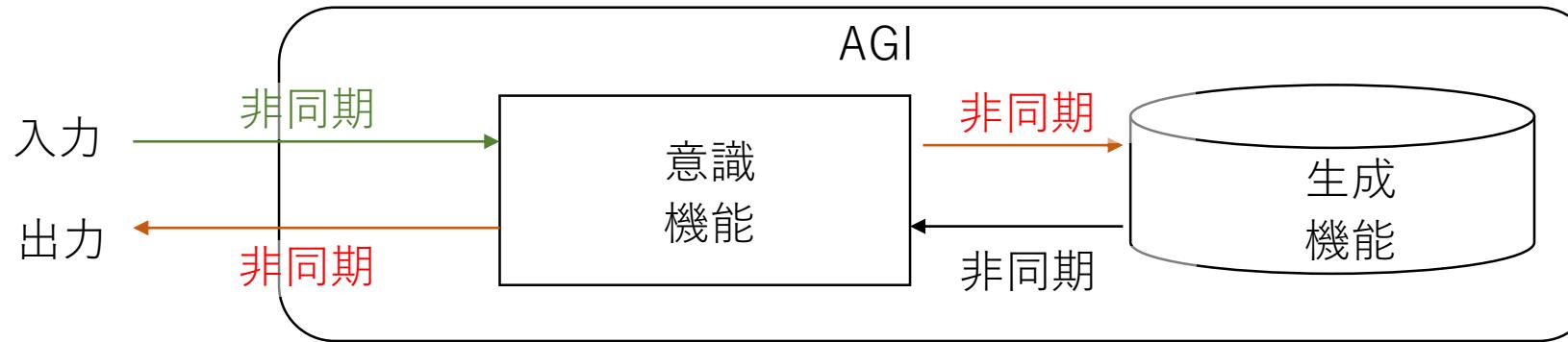
生成機能は、従来の生成AIが応用できます。

意識の機能は、新たに創る必要があります。

3. 生成AIと意識によるAGIの基本モデル

非同期出力

ここからは、AGIの模式図を元に詳細を考えていきましょう。



意識への入力と出力がありますが、これらは非同期です。

意識は、自由なタイミングで、運動神経や生成機能へ出力できます。

何も出力しないというのも、出力内容の選択肢の一つです。

つまり、常時、出力内容を選択しています。

ここからは、AGIの模式図を元に詳細を考えていきましょう。

まずは、意識を見てください。

意識への入力と出力がありますが、これらは非同期です。

意識は、自由なタイミングで、運動神経や生成機能へ出力できます。

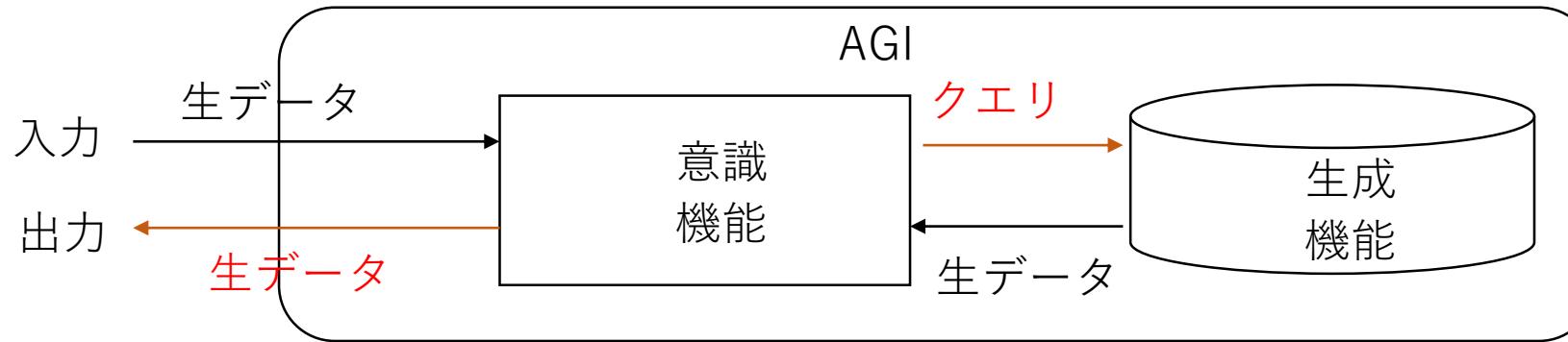
何も出力しないというのも、出力内容の選択肢の一つです。

つまり、常時、出力内容を選択しています。

3. 生成AIと意識によるAGIの基本モデル

出力形式

意識から出力されるデータ形式を見てみましょう。



外部への出力形式は、生データです。

一方で、生成機能への出力形式はクエリです。

意識から出力されるデータ形式を見てみましょう。

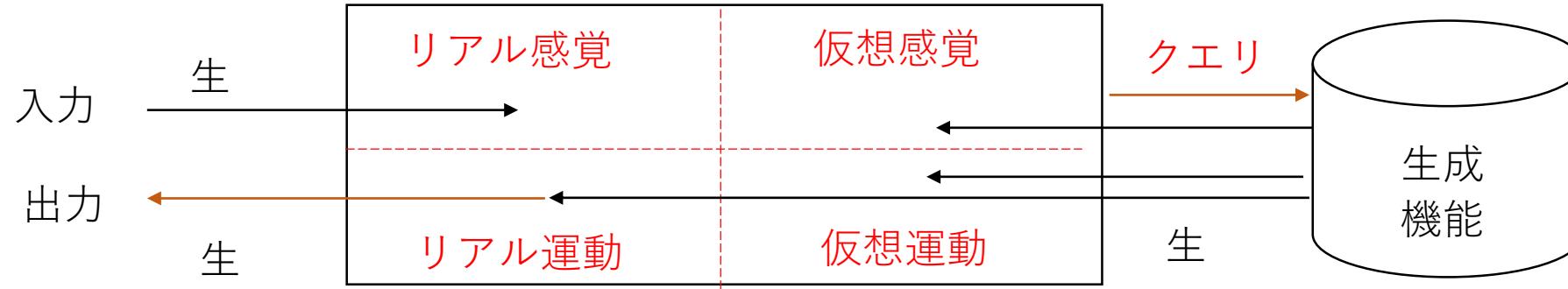
外部への出力形式は、生データです。

一方で、生成機能への出力形式はクエリです。

できれば、出力形式を統一したいですね。

3. 生成AIと意識によるAGIの基本モデル ドメイン

意識を四つのドメインに分けました。



- ・生成機能は、リアル感覚ドメイン以外へ出力できます。
- ・リアル運動ドメインに書き込まれたデータは、自動的に外部出力されます。
- ・外部出力したい場合、生成機能へ発注します。
- ・意識からの出力は、生成機能へのクエリへ統一されました。

意識を四つのドメインに分けました。

生成機能は、リアル感覚ドメイン以外へ出力できます。

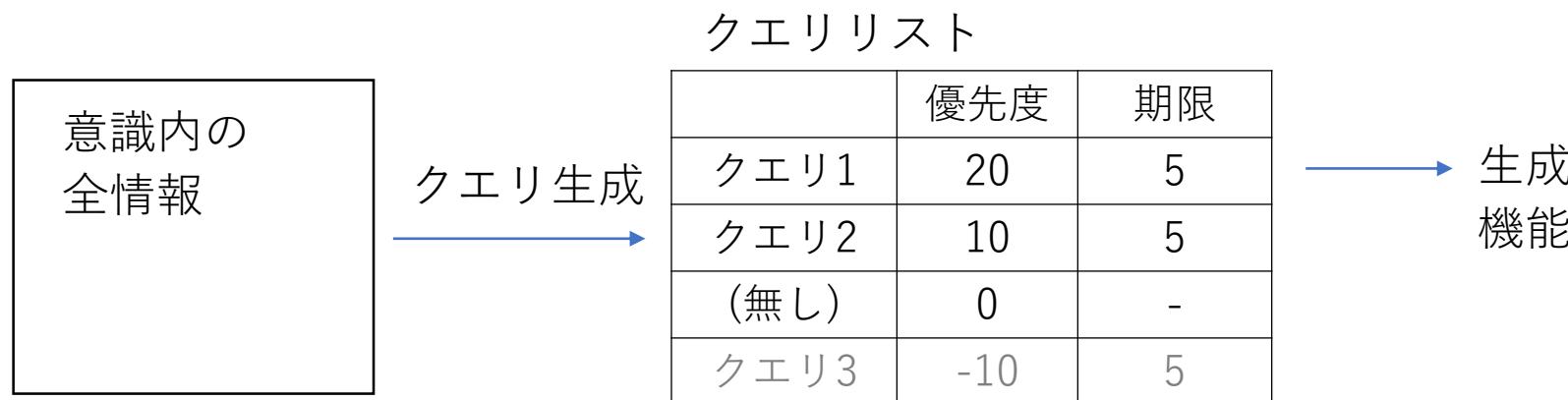
リアル運動ドメインに書き込まれたデータは、自動的に外部出力されます。

外部出力したい場合、生成機能へ発注します。

意識からの出力は、生成機能へのクエリへ統一されました。

3. 生成AIと意識によるAGIの基本モデル

クエリリスト



- ・意識は、意識内の全情報から、クエリを生成し、リストへ追加します。
- ・優先度が高い順に、生成機能へ送られて実行されます。

意識が行う仕事が明確になってきました。

意識は、意識内の全情報から、クエリを生成し、リストへ追加します。

各クエリには、優先度と期限の情報を付与します。

優先度が高い順に、生成機能へ送られて実行されます。

ただし、「無し」という特殊なクエリより優先度が低いものは、処理しません。

3. 生成AIと意識によるAGIの基本モデル クエリの要素

- | | |
|--|----------------|
| ・クエリ内容 | ランダム性が許容される。 |
| ・期限 | 生成AI等で、生成して良い。 |
| ・優先度 = 優先度関数 $f(\text{クエリ内容}, \text{意識内の全情報})$ | |

優先度は、クエリ内容と意識内の全情報の写像です。

ランダム性があってはいけません。

ランダム性が許されると、

都合が良い結果になるまで計算し直すことができてしまうからです。

クエリには、クエリ内容、期限、優先度の三要素があります。

クエリ内容と期限には、ランダム性が許容されます。

そのため、生成AI等で生成して構いません。

優先度は、優先度関数で決まります。

優先度は、クエリ内容と意識内の全情報の写像です。

ランダム性があってはいけません。

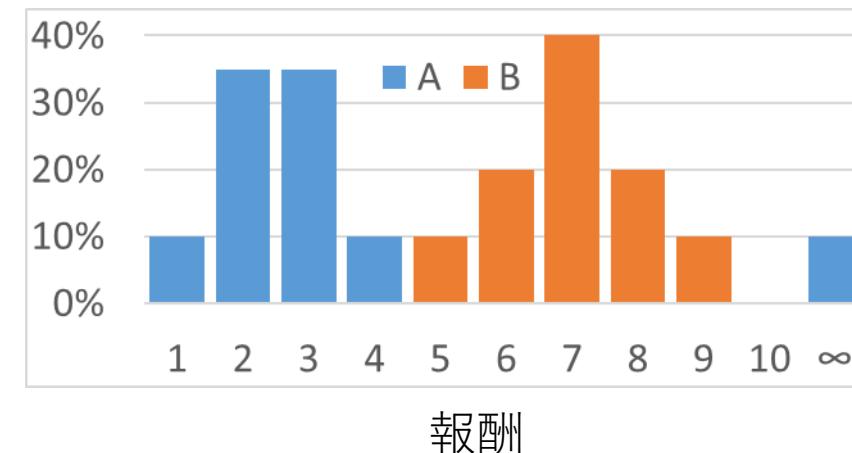
ランダム性が許されると、都合が良い結果になるまで計算し直すことができてしまうからです。

3. 生成AIと意識によるAGIの基本モデル

報酬の分布

優先度 = 未来に受け取る報酬の推測値

報酬は、期待値ではなく、確率分布のまま保持して比較します。



A の期待値 = ∞

B の期待値 = 7

$B > A$ の確率 = 90%

報酬が首位の確率が首位のものを選ぶなら、Bです。

優先度は、未来に受け取る報酬の推測値言い換えることができます。

報酬は、期待値ではなく、確率分布のまま保持して比較します。

例えば、分布Aは、10%の確率で報酬が無限大なので、期待値も無限大です。

分布Bの期待値は7なので、期待値ではAが上位です。

ですが、90%の確率で、分布Bの方が報酬が大きくなります。

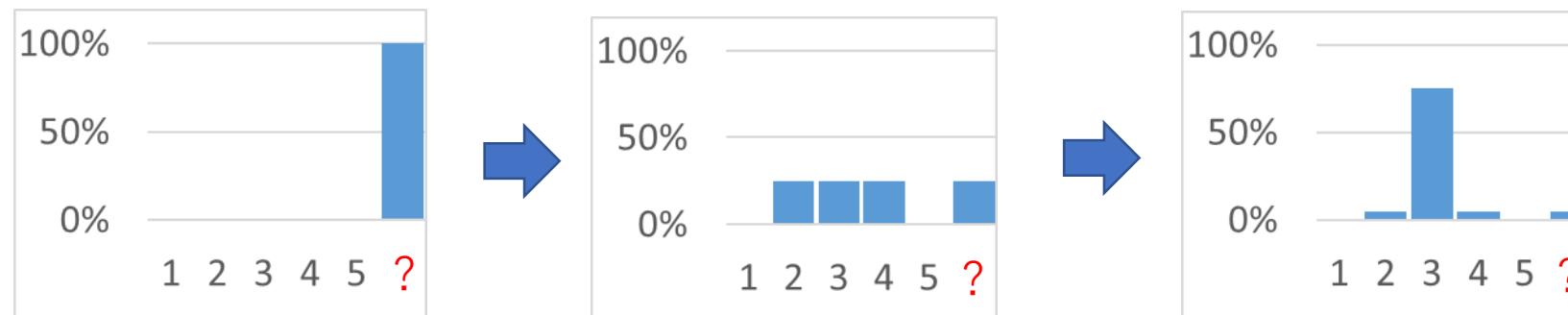
報酬が首位の確率が首位のものを選ぶなら、Bです。

3. 生成AIと意識によるAGIの基本モデル 分布の更新

未来の報酬は帰納推論する必要があります。

帰納推論では、考慮することを増やすほど、精度が良くなります。

フレーム問題が起こるため、全てを考慮することはできません。



そこで、徐々に考慮することを増やして、推測値を更新します。

初期状態では、不明(?)という値が100%の分布になります。

未来の報酬は帰納推論する必要があります。

帰納推論では、考慮することを増やすほど、精度が良くなります。

フレーム問題が起こるため、全てを考慮することはできません。

そこで、徐々に考慮することを増やして、推測値を更新します。

初期状態では、不明(?)という値が100%の分布になります。

3. 生成AIと意識によるAGIの基本モデル 不明の比較

「不明」という値は、どうやって比較するか説明します。

A=2, B=1, C=不明, D=不明

値が首位の確率が首位のものは、どれでしょうか？

BはAより小さいため、首位の確率は0%です。

残りのA,C,Dが首位の確率が、各1/3になります。

一つだけ選びたい場合、首位が複数なら、ランダムに選ぶしかありません。

「不明」という値は、どうやって比較するか説明します。

「A=2、B=1、C=不明、D=不明」だとします。

値が首位の確率が首位のものは、どれでしょうか？

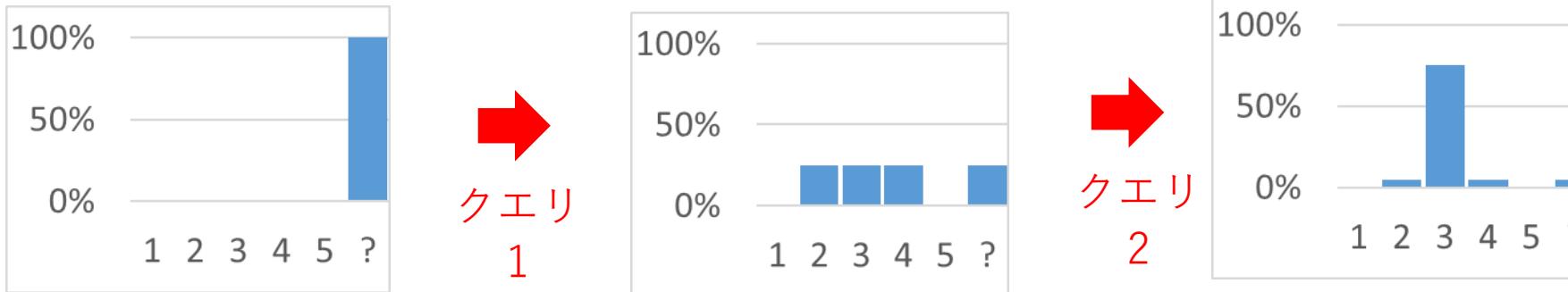
BはAより小さいため、首位の確率は0%です。

残りのA,C,Dが首位の確率が、各1/3になります。

一つだけ選びたい場合、首位が複数なら、ランダムに選ぶしかありません。

3. 生成AIと意識によるAGIの基本モデル 推論クエリ

段階的な帰納推論を、どこまで行うかという優先度を考える必要があります。
そこで、帰納推論という仕事も、クエリリストに追加して管理します。



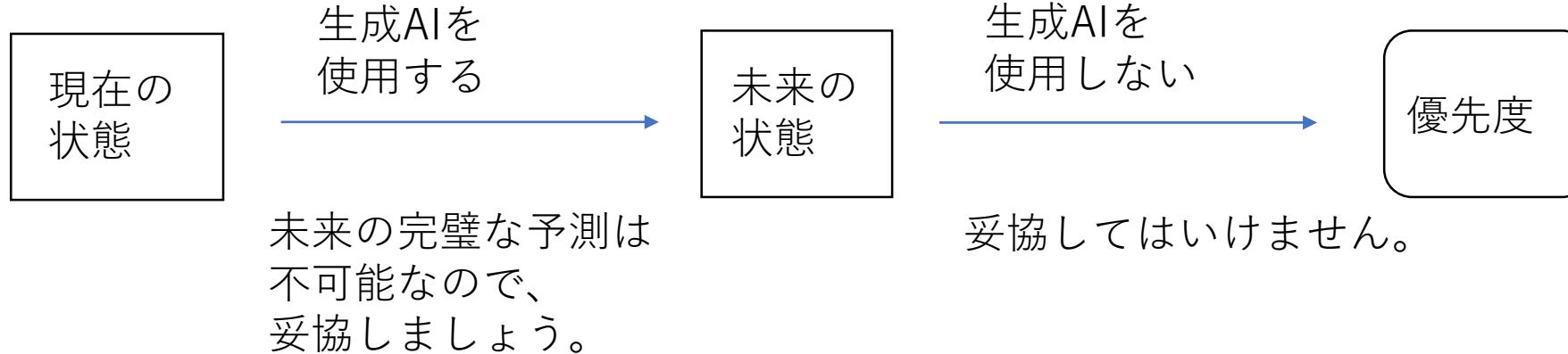
生成機能が帰納推論する形になります。
ただし、優先度関数にはランダム性があってはいけないことを忘れてはいけません。

段階的な帰納推論を、どこまで行うかという優先度を考える必要があります。
そこで、帰納推論という仕事も、クエリリストに追加して管理します。
生成機能が帰納推論する形になります。

3. 生成AIと意識によるAGIの基本モデル

未来の生成

生成機能が優先度を決めるのは許されません。



生成機能が優先度を決めるのは許されません。

生成AIを過信してはいけません。

生成AIには、推測される未来の状態を生成してもらいます。

未来の完璧な予測は不可能なので、妥協しましょう。

ですが、未来の状態から優先度を算定する部分は、妥協してはいけません。

優先度は、生成AIを使わずに算定します。

3. 生成AIと意識によるAGIの基本モデル

報酬と優先度

$$\text{優先度} = \int_{\text{now}}^{+\infty} \text{報酬} dt$$

時間と共に変化する

$$\text{優先度} = \int_{-\infty}^{+\infty} \text{報酬} dt$$

時間と共に変化しない

- ・大小関係だけが問題なので、共通する過去の部分は無視できる。
- ・同じ精度で推測できているのなら、
近い未来の報酬も、遠い未来の報酬も同じ価値です。

報酬と優先度の関係を数式で表しました。

典型的には、現在から無限の未来までに受け取る報酬の総和を優先度とします。

ですが、この場合、時間と共に変化していまいります。

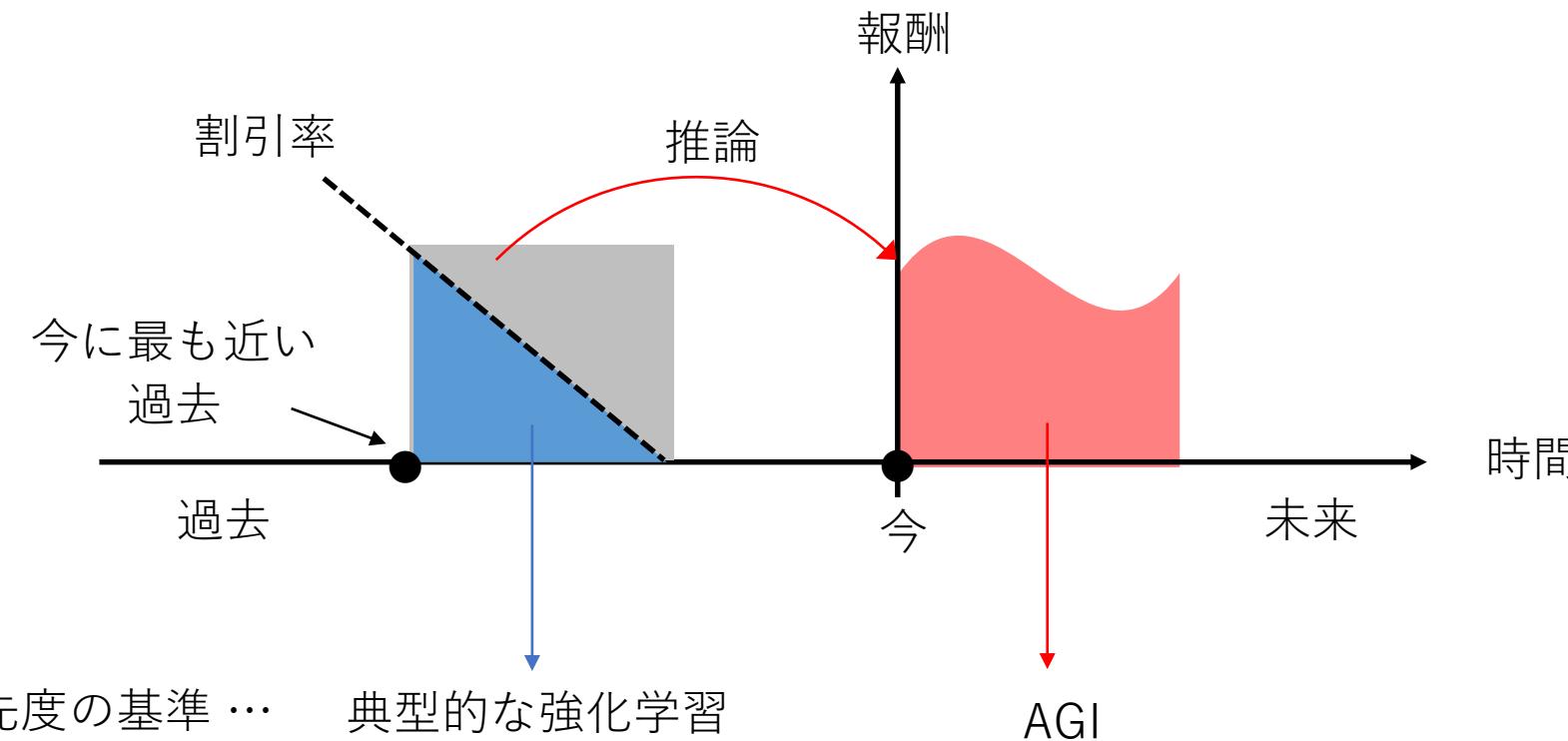
そこで、無限の過去から無限の未来の総和にすれば、時間と共に変化しません。

大小関係だけが問題なので、共通する過去の部分は無視できます。

同じ精度で推測できているのなら、近い未来の報酬も、遠い未来の報酬も同じ価値です。

3. 生成AIと意識によるAGIの基本モデル

強化学習



典型的な強化学習とAGIにおける報酬の違いを図示しました。

強化学習では、過去の報酬が優先度の基準です。

似た過去を基準にして、割引率を掛けています。

AGIでは、未来の報酬が優先度の基準です。

未来の報酬は推論します。

3. 生成AIと意識によるAGIの基本モデル

無効な報酬設定

報酬: 目的に合わせてプログラマーが自由に設定

ただし、学習前に設定できる必要があります。

例えば、「質問者が喜ぶこと」を報酬に設定することはできません。

なぜなら、学習前は「質問者」や「喜ぶ」といった概念を知らないからです。
メモリ上のアドレスが不明な変数は比較できません。

報酬は、目的に合わせてプログラマーが自由に設定します。

ただし、学習前に設定できる必要があります。

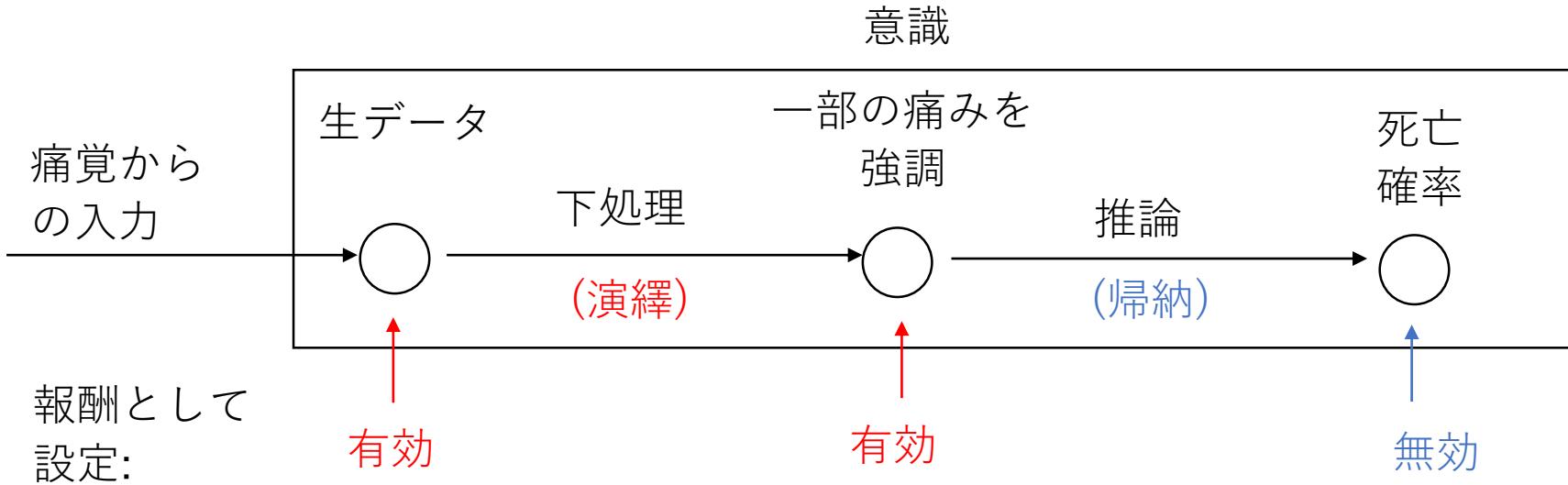
例えば、「質問者が喜ぶこと」を報酬に設定することはできません。

なぜなら、学習前は「質問者」や「喜ぶ」といった概念を知らないからです。

メモリ上のアドレスが不明な変数は比較できません。

3. 生成AIと意識によるAGIの基本モデル

有効な報酬設定



生データによって必然的に決まる写像であることが、有効な報酬の条件です。

感覚神経からの入力された生のデータなら、報酬の設定に使えます。

例えば、全ての痛みの合計は、有効な報酬設定です。

演繹で計算できる下処理後のデータでも有効です。

さらに、「痛み」からは、「死亡確率」を推測できます。

ですが、この「死亡確率」を報酬に設定するのは無効です。

恣意的な帰納推測することが許されてしまうからです。

生データによって必然的に決まる写像であることが、有効な報酬の条件です。

3. 生成AIと意識によるAGIの基本モデル 帰納と演繹

報酬の計算



推論タイプ	帰納	演繹
計算量	考慮するフレームを 拡げるほど 無限に増大	有限
中止	する	しない
クエリリスト	入れる	入れない
計算者	生成機能	意識機能

報酬は、既知の情報だけから演繹推論で計算します。

帰納推論では、考慮するフレームを拡げるほど計算量が無限に増大します。

優先度を決めて、適宜、計算を打ち切る必要があります。

そのため、クエリリストへ入れて、生成機能が処理します。

一方、演繹推論では、考慮するべきことが有限なので、計算量も有限です。

報酬の計算が不正確だとバイアスやハルシネーションを生みます。

計算量が有限なので、打ち切らずに計算します。

報酬の計算はクエリリストに入れず、生成機能ではなく、意識機能で行います。

3. 生成AIと意識によるAGIの基本モデル

クエリの追加

生成機能		意識機能	
内容	優先度	期限	優先度の時間変化
クエリ1	20	5	
クエリ2	10	5	
クエリ追加	1	-	
(無し)	0	-	
クエリ3	-10	5	

初期状態から、「クエリ追加」という特別なクエリが存在します。

生成機能は、現在のクエリリストを参考に、新しいクエリを追加します。

不要になったクエリの削除や置換も行います。

クエリリストの説明に戻ります。

クエリの内容は、生成機能が決めます。

クエリの優先度と期限は、意識機能が決めます。

期限というのは、優先度の時間変化です。

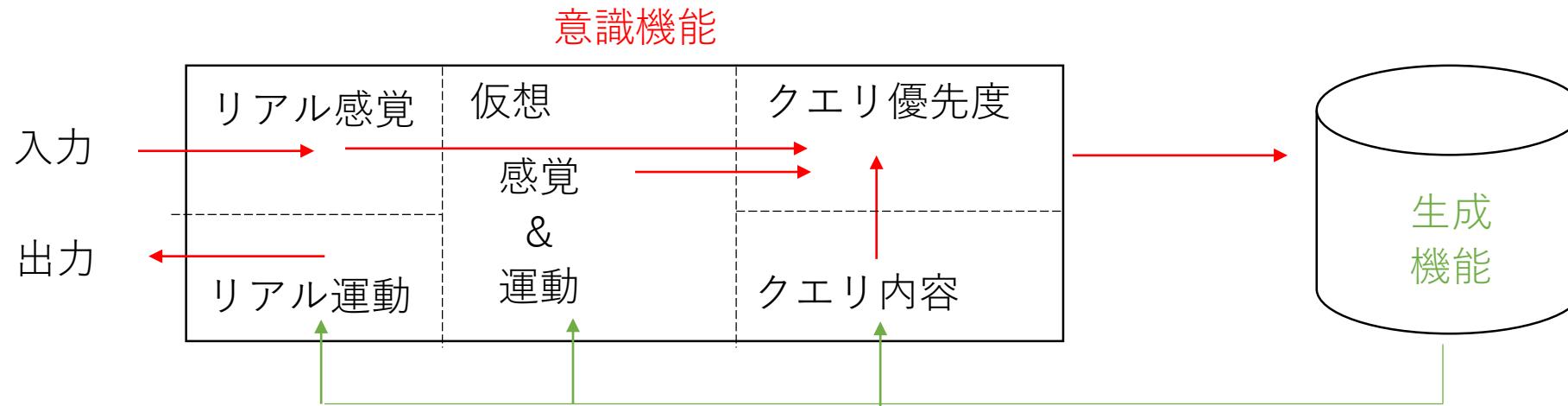
初期状態から、「クエリ追加」という特別なクエリが存在します。

生成機能は、現在のクエリリストを参考に、新しいクエリを追加します。

不要になったクエリの削除や置換も行います。

3. 生成AIと意識によるAGIの基本モデル

逐次処理AGI



逐次処理する場合、意識機能と生成機能が、交互に働きます。

ただし、一定時間経過や入力イベントで割り込みすることもできます。

AGIの模式図にクエリ関連を追加しました。

逐次処理する場合、意識機能と生成機能が、交互に働きます。

ただし、一定時間経過や入力イベントで割り込みすることもできます。

3. 生成AIと意識によるAGIの基本モデル

生成機能の形式

生成機能

出力 文字列 = $f(\text{文字列})$ 入力

入出力が一つの文字列変数である必要ありません。

意識へ書き込む内容 = $f(\text{処理タイプ}, \text{処理内容}, \dots)$

処理のタイプによって、使用するAIを切り替えることも可能です。

また、生成AIである必要もありません。

次に、生成機能の方を考えていきましょう。

生成AIを使用可能ですが、チャットボット形式である必要はありません。

すなわち、入出力が一つの文字列変数である必要ありません。

例えば、一つ目の引数で、処理タイプを指定します。

二つ目以降の引数で、処理内容を指定します。

出力は、意識へ書き込む内容になります。

処理のタイプによって、使用するAIを切り替えることも可能です。

また、生成AIである必要もありません。

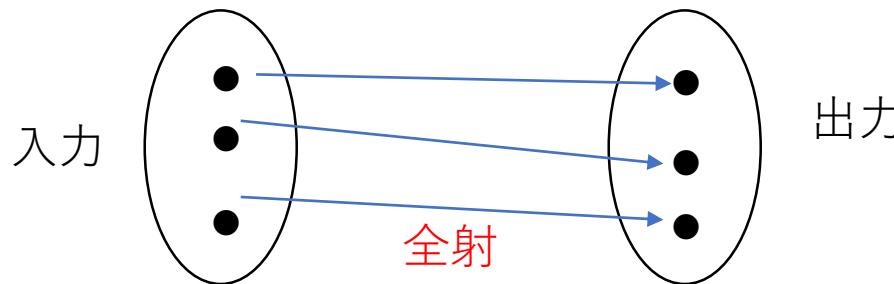
3. 生成AIと意識によるAGIの基本モデル

全射

生成機能には必須条件があります。

出力は、入力に対する写像です。

この写像は、**全射**でなければなりません。



全てのパターンを出力できる入力が存在していなければいけません。

それは、どんなものでも生成可能だということです。

ランダムさを持たせれば、この条件を満たすのは容易です。

生成機能には必須条件があります。

出力は、入力に対する写像です。

この写像は、全射でなければなりません。

例えば、出力は1000ビットだとすると、2の1000乗のパターンがあります。

全てのパターンを出力できる入力が存在していなければいけません。

それは、どんなものでも生成可能だということです。

ランダムさを持たせれば、この条件を満たすのは容易です。

3. 生成AIと意識によるAGIの基本モデル ランダム出力

極論、生成機能による出力の生成は完全なにランダムでも構いません。
チャットボットを例にすると、ランダムな文字列を生成します。

- Qwd wiuf awaiji.
 - Fosi apipjw ejw.
 - Kisv oa cvrnsb?
 - May I help you? ←
- ほとんどは、意識機能によって、
超低評価され棄却されます。
- 極低確率で、偶然に高評価な文章が
できあがります。

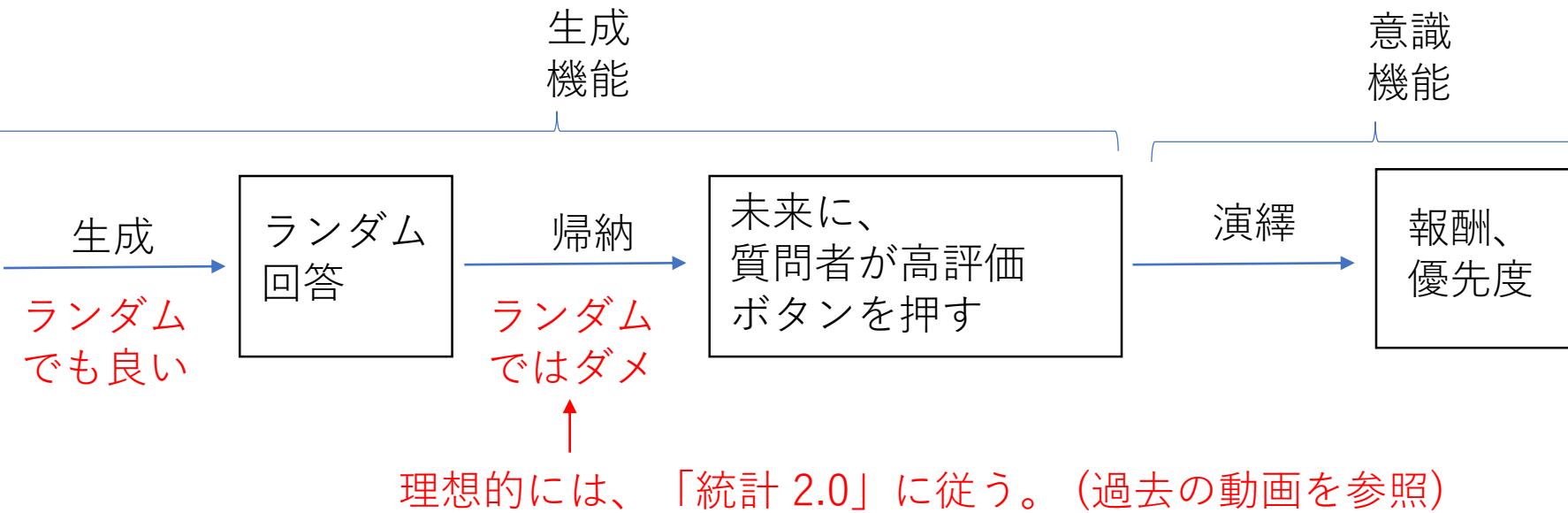
計算能力が無限大なら、必ずベストな回答ができます。
ただ計算効率が悪いだけです。

ですが、生成機能の全てにランダムが許されるのではありません。

極論、生成機能による出力の生成は完全なにランダムでも構いません。
チャットボットを例にすると、ランダムな文字列を生成します。
そのほとんどは、意識機能によって、超低評価され棄却されます。
ですが、極低確率で、偶然に高評価な文章ができあがります。
計算能力が無限大なら、必ずベストな回答ができます。
ただ計算効率が悪いだけです。
ですが、生成機能の全てにランダムが許されるのではありません。

3. 生成AIと意識によるAGIの基本モデル

帰納推論



全ての帰納推論が統計2.0に従っていれば、必ず最適解へ到達します。

回答の生成はランダムでも許されます。

次に、未来に質問者が高評価ボタンを押すと帰納推論します。

帰納推論の部分はランダムではいけません。

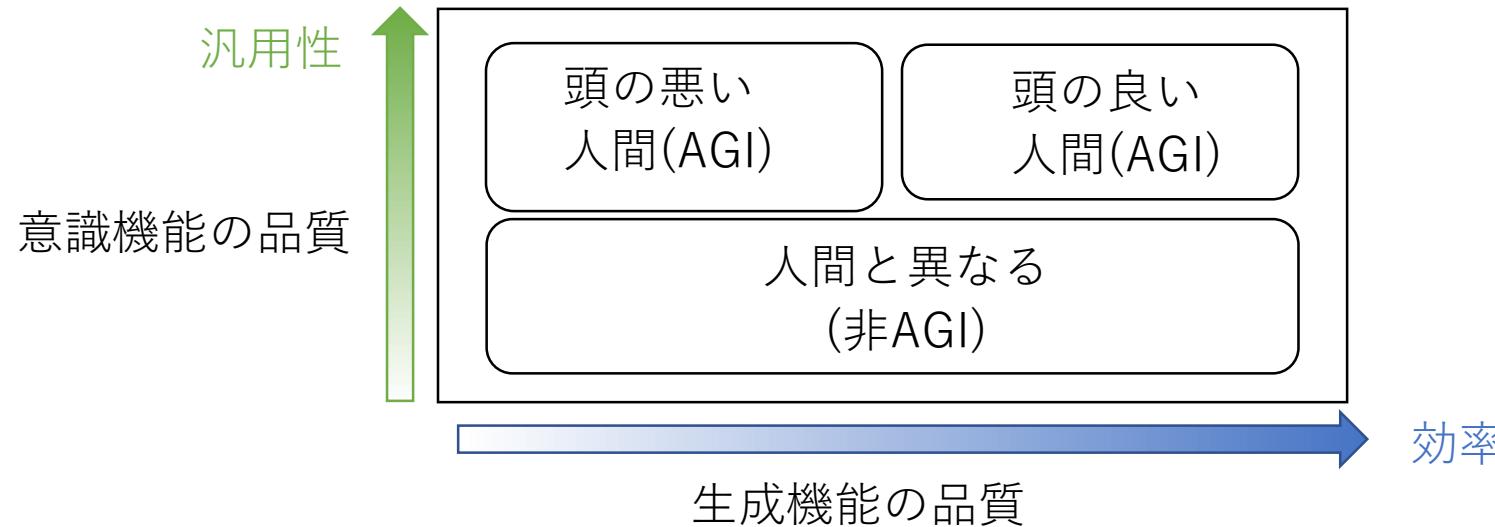
理想的には、統計2.0に従います。

統計2.0については、過去の動画を参照してください。

全ての帰納推論が統計2.0に従っていれば、必ず最適解へ到達します。

3. 生成AIと意識によるAGIの基本モデル

汎用性と効率



効率が悪くても、汎用性があれば、頭が悪い人間と同等です。

効率は、ハードウェアで改善できます。

効率の良し悪しと、汎用性の良し悪しは無関係です。

知能は、汎用性と効率の分けて考えることができます。

意識機能の性能が良いほど、汎用性が高まります。

生成機能の性能が良いほど、効率が高まります。

効率が悪くても、汎用性があれば、頭が悪い人間と同等です。

効率は、ハードウェアで改善できます。

効率の良し悪しと、汎用性の良し悪しは無関係です。

3. 生成AIと意識によるAGIの基本モデル

AGIと脳の比較

	このAGI	脳
処理	逐次	並列
意識機能と生成機能	分離	融合
品質	限界なし	生体的限界

脳は生体なので品質に限界があります。

一方で、このAGIは、品質に限界がないのが利点です。

脳を参考に、僅かに品質を犠牲にして、高速化する余地があります。

このAGIと脳を比較してみましょう。

逐次処理と並列処理の違いがあります。

意識機能と生成機能が分離しているか融合しているかも異なります。

また、脳は生体なので品質に限界があります。

一方で、このAGIは、品質に限界がないのが利点です。

脳を参考に、僅かに品質を犠牲にして、高速化する余地があります。

3. 生成AIと意識によるAGIの基本モデル まとめ

意識機能と生成機能から成るAGIの基本モデルができました。

まだ出来ていないこと

1. 具体的な生成機能
2. 効率化

この二つを妥協しても汎用性は達成できます。

また、これらは目的や装置に依存します。

ですが、実用化するには必要なことです。

まとめです。

意識機能と生成機能から成るAGIの基本モデルができました。

まだ出来ていないことが二つあります。

1. 具体的な生成機能
2. 効率化

この二つを妥協しても汎用性は達成できます。

また、これらは目的や装置に依存します。

ですが、実用化するには必要なことです。

生成AIと意識を活用したAGI

あとがき

最後に、この動画をご覧の皆様へお願ひです。

私には考える力しかありません。

お金、時間、設備、データ、人材、知名度など、全てありません。

貴方の力が必要です。

この動画を見ているとき、私はもうこの世にいないかもしれません。

どうか、AGIを完成させてください。

それだけが私の願いです。

お問い合わせ: <https://ultagi.org/>

最後に、この動画をご覧の皆様へお願ひです。

私には考える力しかありません。

お金、時間、設備、データ、人材、知名度など、全てありません。

貴方の力が必要です。

この動画を見ているとき、私はもうこの世にいないかもしれません。

どうか、AGIを完成させてください。

それだけが私の願いです。

お問い合わせ先

お問い合わせは、
こちらからお願いします。

<https://ultagi.org/>