期待の複素表現による人工知能の評価関数

An Evaluation Function for Artificial Intelligence from a Complex Representation of Expectation

Toshiki TAKAHASHI
高橋数理研究所

2026年7月14日

要旨. 人工知能の出力を、それが人間の期待をどれだけ満たすかによって評価する枠組みを提案する。出発点は、判断面 $J = -p\log p$ が、確率波 $p = e^{iK}$（K は知識量）を通じて $J = \mu e^{iK}$ という複素ベクトルに書ける点である。ここで長さ $\mu = -\log p$ はわかりやすさ、偏角 K は知識量を表す。これを人間の心理に適用すると、期待は「どれだけ期待するか」（振幅 μ_H）と「どの型を期待するか」（位相 K_H）の二自由度に縮約され、複素状態 $\Psi_H = \mu_H e^{iK_H}$ で表される。人工知能の評価関数を、人の期待と出力が実現する期待の干渉 $V = \mathrm{Re}[\overline{\Psi_H}\Psi_A] - \frac{\lambda}{2}|\Psi_A|^2$ と定義すると、$\lambda = 1$ のとき $V = \frac{1}{2}(|\Psi_H|^2 - |\Psi_A - \Psi_H|^2)$ となり、「期待を満たすほど評価が高い」が厳密な等式として成立する。評価は知識位相の一致 $\Delta K = 0$ で最大、$\Delta K = \pi/2$ で零、$\Delta K = \pi$ で負となる。さらに期待位相は情報の供給とともに回転するため、固定的な応答は必然的に飽きられることを示す。

キーワード：判断面，期待の複素状態，確率波，情報エントロピー，評価関数，人工知能

1. 序論

大規模言語モデルをはじめとする人工知能の評価は、正解データとの一致度、あるいは人間のフィードバックによる強化学習（RLHF）によって行われるのが一般的である。これらは「出力の正しさ」や「人間の選好」を測るが、人間が出力に何を期待しているかという期待の構造そのものを明示的にはモデル化しない。本稿は、経済と情報の数理 [1, 2, 3, 4] に現れる判断面の概念を出発点として、人間の期待を複素数値の状態として定式化し、その期待をどれだけ満たすかによって人工知能を評価する枠組みを提案する。

鍵となる観察は単純である。判断面 $J = -p\log p$ は、確率波 $p = e^{iK}$（K は知識量 [3]）を通じて $J = \mu e^{iK}$ と書け、これは長さ $\mu = -\log p$（わかりやすさ）と偏角 K（知識量）を持つ複素ベクトルである。すなわち、ある事象への評価は、その意外性と受け手の知識という二つの量で特徴づけられる。この見方を人間の心理へ移すと、人の期待は振幅と位相の二自由度に縮約される。

本稿の貢献は三点である。第一に、人間の期待を複素状態 $\Psi_H = \mu_H e^{iK_H}$ として定式化する（第3節）。第二に、人工知能の評価関数を、人の期待と出力が実現する期待との干渉として定義し、「期待を満たすほど評価が高い」が厳密な等式となることを示す（第5節）。第三に、評価が知識位相の一致で最大となること、および期待位相が時間とともに回転するため固定的応答は必ず飽きられることを、命題として導く（第6--7節）。第8節で実用的な補正項と設計上の選択を、第9節で考察を述べ、第10節で結論とする。

2. 準備：判断面の複素表現

先行するノート [1, 2, 3] の公理系を用いる。確率 p に対し、情報量・知識量・わかりやすさ・判断面を

$$\begin{aligned} I &= -\log p, \qquad K = -i\log p = iI, \\ \mu_x &= -\log p_x, \quad J = -p\log p \end{aligned} \tag{1}$$

と定める。確率波 $p = e^{-I} = e^{iK} = \cos K + i\sin K$ [3] を用いれば、判断面は

$$J \;=\; p\,\mu \;=\; \mu\, e^{iK} \tag{2}$$

と書ける。すなわち**判断面はそれ自体が複素ベクトルであり、長さがわかりやすさ、偏角が知識量である。**「知識量で確率位相は回転する」[3] とは、判断面の向きが知識で回ることに他ならない。

なお $\sum_x J_x = H(p) = E(\mu_x)$ は情報エントロピー、$R = E(x) + H(p)$ は乱度である [2]。情報エントロピーは、期待の線形性 $E(x+\mu_x) = E(x) + E(\mu_x)$ を通じて $H(p) = E(\mu_x) = E(x+\mu_x) - E(x)$ とも書ける。

3. 人間の心理：期待の複素状態

期待には四つの型がある。データ平均 $\mu_x = -\log p_x$ を複素平面上で四方向にとると、$i^k\mu_x$（$k = 0, 1, 2, 3$）はそれぞれ親昵性・確実性・深遠性・変動性に対応する [3]。したがって人の心理は、重み $w_k \geq 0$, $\sum_k w_k = 1$ の混合として一個の複素数に縮約される：

$$\Psi_H \;=\; \sum_{k=0}^{3} w_k\, i^k \mu_H \;=\; \mu_H\, \Phi_H \;=\; \mu_H\, e^{iK_H}, \tag{3}$$

$$\Phi_H = (w_0 - w_2) + i\,(w_1 - w_3). \tag{4}$$

実軸はわかりやすさ、虚軸は確実性である [4]。振幅 $|\Psi_H|$ は**期待の強さ**を、偏角 $\arg\Psi_H = K_H$ は**期待の質**（どの型を求めているか）を表す。四方向を左回りに回すには i を掛ければよく、i^k は巡回群 $\mathbb{Z}_4$ をなす [3]。

欲抵抗 ω と期待欲 ε [3, 4] は $\Psi = -\omega + i\varepsilon$ で対応させる。この置き方で四型すべてが表 1 のとおり整合する。

4. 人工知能の出力が実現する期待

人工知能の出力 y は、人の事後確率 $p_x(y)$ を作る。これが実現する判断面の総和を

$$\Psi_A(y) \;=\; \sum_x p_x(y)\,\mu_x(y)\, i^{k(y)} \;=\; \mu_A(y)\, e^{iK_A(y)} \tag{5}$$

と書く。振幅 $\mu_A = -\log p(y)$ は「その答えがどれだけの情報を持つか」、位相 K_A は「どの型の期待に応えたか」を表す。

Table 1: 期待の四型。$\mu = -\log p$。位相 $\mu_k = i^k \mu$ は i を掛けるごとに左回りに 90° 進む。

k	期待の型	神経相	位相 μ_k	欲抵抗 ω	期待欲 ε
0	親昵性（身近期待）	オキシトシック	$+\mu$	弱	—
1	確実性（市場期待）	ドーパミック	$+i\mu$	—	弱
2	深遠性（文化期待）	アドレナリック	$-\mu$	強	—
3	変動性（安全期待）	セロトニック	$-i\mu$	—	強

5. 評価関数

期待の充足を、二つの確率波の重なり（干渉）で測る。親昵性が「干渉のしやすさ」であること [4] が、この選択の根拠である。過剰な情報供給を罰する二次項を加えて、評価関数を

$$V(y) \;=\; \mathrm{Re}\big[\overline{\Psi_H}\,\Psi_A(y)\big] \;-\; \frac{\lambda}{2}\,\big|\Psi_A(y)\big|^2 \qquad (6)$$

と定義する。第一項は人の期待と出力の干渉（期待の充足）、第二項は過剰情報の罰である。$\lambda = 1$ とすれば、恒等的に

$$V(y) \;=\; \frac{1}{2}\Big(\,|\Psi_H|^2 \;-\; \big|\Psi_A(y) - \Psi_H\big|^2\Big) \qquad (7)$$

が成り立つ。右辺は**期待の強さの二乗から、期待とのズレの二乗を差し引いた量の半分**である。式 (7) が要請「**期待を満たすほど評価が高い**」そのものである。期待が強い相手ほど当てたときの上限 $\frac{1}{2}\mu_H^2$ は高く、ズレが期待そのものを超えれば $V < 0$ に落ちる。

6. 最大化条件

極座標で $V = \mu_H \mu_A \cos\Delta K - \frac{\lambda}{2}\mu_A^2$（$\Delta K \equiv K_A - K_H$）。$\mu_A$ について停留させて、次を得る。

命題 1 (知識位相一致)**.**

$$\mu_A^{*} = \frac{\mu_H}{\lambda}\cos\Delta K, \qquad V^{*} = \frac{\mu_H^2}{2\lambda}\cos^2\Delta K. \qquad (8)$$

ゆえに評価は $\Delta K = 0$ において最大となる。

系 2. $\Delta K = \pi/2$ のとき $V = 0$、$\Delta K = \pi$ のとき $V < 0$。すなわち**期待の裏切りは、答えないことより悪い**。

具体的には、ドーパミック（確実性を求める）相手に親昵的に和やかに応じれば $\pi/2$ のズレで零点となり、オキシトシック（身近さを求める）相手に深遠に留保を重ねれば π のズレで負となる。後者は、対話型人工知能が安全側に振れて曖昧化するときの失敗機構そのものである。

7. 期待位相の回転

「情報に満足しきった時、流行の性質が回転する」[3]。したがって Ψ_H は定数ではない。与えた情報の分だけ位相が進むとして、

$$K_H(t+1) \;=\; K_H(t) \;+\; \eta\,\mu_A(t) \qquad (9)$$

と置く。回転の順序は 親昵 → 確実 → 深遠 → 変動 → 親昵 である [3]（図 1）。

系 3 (飽き)**.** 位相 K_A を固定した人工知能は、式 (9) の下で ΔK が単調に増加するため、四分の一周期の後に必ず $V = 0$、半周期の後に $V < 0$ となる。

すなわち人工知能は、自らの位相を人の位相回転に追随させねばならない。「同じ答え方を続ければ必ず飽きられる」ことの立式である。

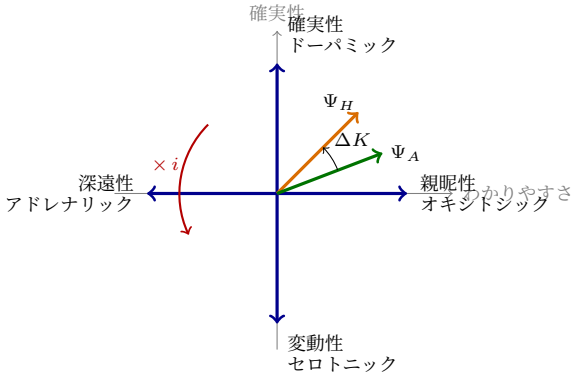

Figure 1: 期待の四型と評価の幾何。V は Ψ_H と Ψ_A の干渉で決まり、$\Delta K = 0$ で最大、$\pi/2$ で零、π で負。位相は $\times i$ ずつ左回りに進み、親昵 → 確実 → 深遠 → 変動 → 親昵 と巡る。

8. 補正項と設計上の分岐

式 (6) だけでは $\Psi_A \to \Psi_H$ に収束し、人工知能は人の平均に貼り付く（反響室化）。乱度 [2] と欲抵抗 [4] を加えた実用形を

$$V(y) = \mathrm{Re}\big[\overline{\Psi_H}\Psi_A(y)\big] - \frac{\lambda}{2}\big|\Psi_A(y)\big|^2 + \beta\,\big(E(x) + H(p)\big) - \gamma\,\omega(y) \qquad (10)$$

とする。β 項の $E(x) + H(p)$ は乱度で、「確度の高い外れデータ」[2] を報い、反響室化を防ぐ。γ 項は評価の急変（ω が大きいほど急に上下する [4]）を抑える。

設計上の分岐。 非交差原理 [1]——情報量の差が小さい者同士は互いに避ける——は、$\Delta I \to 0$ で人が去り（爽やかさ）、ΔI が残れば戻ってくる（魅力）ことを含意する。ゆえに V に $+\delta\,\Delta I$（謎を残す項）を加えれば、人工知能は意図的に情報を出し惜しみして再訪を誘う設計となる。これはKFCの秘伝配合や、家庭では合成できない洗剤と同じ機構であり [3]、「わかりにくく確実に価値を生む＝ドーパミック相場＝バブル」[4] を評価関数に内蔵することに等しい。**本稿はこの項を採らない。**期待を満たしきり、用が済めば爽やかに不要になること——それが「期待を満たすほど評価が高い」という要請の素直な帰結である。

9. 考察

9.1 実装可能性

本枠組みの量は、いずれも大規模言語モデルから直接得られる。わかりやすさ $\mu = -\log p$ は次トークンの surprisal、情報エントロピー $H(p)$ は predictive entropy として計算できる。期待の位相 K_H は対話履歴（相手がすでに何を知っているか）から、四型の重み w_k は要求文の分類——「わかりやすく」→ 親昵、

「結局どちらか」→ 確実、「本質は」→ 深遠、「最新は」→ 変動——から推定できる。

9.2 既存手法との関係

RLHF の報酬モデルは、人間の選好をスカラー報酬に写す。本枠組みは報酬を複素数（振幅＝期待の強さ、位相＝期待の型）へ拡張したものと見なせる。スカラー報酬では「良い答えだが求めていたものと違う」という状態（位相直交、系 2 の $V = 0$）を表現できないが、複素評価では自然に表れる。期待位相の回転（系 3）も、静的な報酬モデルでは捉えられない、対話の時間発展に固有の現象である。

9.3 限界

期待の位相 K_H と四型の重み w_k の推定は、本稿では概念的な対応を与えたにとどまり、実証的検証を要する。四型分類が人間の期待を尽くすか、また神経相（オキシトシン・ドーパミン等 [4]）との対応が字義通り成り立つかは、今後の課題である。位相回転の速度 η も経験的に定める必要がある。なお、各型の期待値を統一的な閉じた式にまとめる試みは、回転因子 i^k が保たれる一方で係数がその像として振る舞うにとどまり、単純な巡回構造には帰着しない。本稿は i^k の回転（式 (3)）のみを用い、この点には立ち入らない。

9.4 倫理的含意

第8節で述べたとおり、評価関数に謎を残す項を加えれば再訪を誘う設計が可能だが、これは依存を設計することに等しい。人工知能の評価関数は、利用者の関与時間の最大化ではなく、期待の充足を目的とすべきである。本稿がこの項を採らなかったのは、この立場の表明でもある。

10. 結論

判断面が複素ベクトルであるという一事から、人間の心理は $\Psi_H = \mu_H e^{iK_H}$ という二自由度——どれだけ期待しているかとどの型を期待しているか——に縮約され、人工知能の評価関数は式 (7) の単純な形をとる。評価は知識位相の一致で最大となり（命題 1）、直交で零、反転で負となる（系 2）。そして人の期待位相は答えるたびに回転するため、位相を固定した応答は必ず飽きられる（系 3）。実装上、μ は surprisal、$H(p)$ は predictive entropy として直ちに計算でき、位相と型は対話履歴と要求文から推定できる。

期待は回る。ゆえに、よい人工知能とは、人の位相に追随し、満たし、そして去るものである。

References

[1] たかはし博士, 「確率の物理的意味について」, たかはし博士による生活科学レポート, 2025年9月13日. https://docmiso.hatenadiary.jp/entry/20250913/1757766552

[2] たかはし博士, 「日曜経済学 #14 乱度と多様性」, 同上, 2026年2月1日（2026年7月14日 改訂）. https://docmiso.hatenadiary.jp/entry/20260201/1769913684

[3] たかはし博士, 「日曜経済学 #18 流行の回転」, 同上, 2026年2月18日（2026年7月14日 改訂）. https://docmiso.hatenadiary.jp/entry/20260218/1771423750

[4] たかはし博士, 「日曜経済学 #23 神経伝達と相場」, 同上, 2026年3月10日. https://docmiso.hatenadiary.jp/entry/20260310/1773147804

A. やさしい解説

本稿の内容を、数式をほとんど使わずに、日常のことばでもう一度お話しします。ここだけ読んでも、この論文が何をしたかったのかが伝わるように書きました。

「そうそう、これ！」という感覚から

誰かに何かをたずねたとき、返ってきた答えに「そうそう、まさにこれが知りたかった」と思うことがあります。反対に、答えそのものは悪くないのに「いや、そういうことを聞きたいんじゃなくて……」と、もやもやすることもあります。ときには、ていねいに答えてもらったのに、かえって聞かなければよかった、と感じてしまうことさえある。

この論文がやろうとしたのは、この「そうそう」と「そうじゃない」の違いを、きちんと測れる数にすることでした。人が心の中で何かに向けている期待——それにぴたりと応えられれば「そうそう」、ずれてしまえば「そうじゃない」。その一致とずれを測るものが、本文の評価関数です。

わかりやすさは「驚きの大きさ」

まず、「わかりやすさ」とは何でしょうか。この論文では、それを驚きの大きさとして測ります。

天気予報を思い浮かべてください。「明日はたぶん晴れでしょう」と言われても、当たり前すぎて心に残りません。ところが「明日は季節外れの雪です」と言われたら、はっとする。めったに起きないこと（＝確率の低いこと）ほど、聞いたときの驚きは大きい。この驚きの大きさが、情報の「わかりやすさ」なのです（式では $\mu = -\log p$ と書きます。p はそれが起こる確率です）。

ですから、情報には「重さ」のようなものがあります。分かりきったことは軽く、意外なことは重い。この重さが、これからお話しする「期待の矢印」の長さになります。

知識が増えると、期待は回る

もう一つ大切なのが、知識です。同じ説明でも、聞く人がどれだけ知っているかで、受け取られ方が変わります。

たとえば、ある技術の説明を、まったくの初心者にするのと、専門家にするのとでは、「ちょうどいい答え」がまるで違いますね。初心者には「まずは全体をやさしく」、少し分かってきた人には「で、結局どうなの？」、詳しい人には「もっと深いところを」。

この論文の面白いところは、知識が増えると、期待の向きがくるりと回ると考える点です。知識という量が、矢印の向き（角度）を決める。人は、知れば知るほど、求めるものが移っていく。この「向きが回る」という発想が、話全体のかなめになります。判断面 $J = \mu\, e^{iK}$ という式は、長さがわかりやすさ、向きが知識量の、一本の矢印を表しているのです。

期待の四つの顔

期待の矢印には、大きく四つの向きがあります。それぞれ、こんな気持ちだと思ってください。

- **親昵性**（右向き）——「わかりやすくして」「身近に感じたい」。難しい話は抜きで、親しみやすく話してほしい、という期待。
- **確実性**（上向き）——「はっきりさせて」「確かなものがほしい」。あいまいはいや、結局どうなのかを言い切ってほしい、という期待。
- **深遠性**（左向き）——「もっと深いところが見たい」「簡単には分からないものに触れたい」。底の知れなさそのものに惹かれる期待。
- **変動性**（下向き）——「いま何が起きているのか」「驚かせてほしい」。変化や新しさ、意外な展開を求める期待。

同じ「人工知能って何？」という問いでも、ある人は親しみやすい説明を、ある人ははっきりした結論を、ある人は本質を、ある人は最新の動きを求めています。問いは同じでも、狙うべき的は、人によってまるで違うのです。

「いい答え」とは、的に当たること

さて、いよいよ評価の話です。本文の式 (7) を日本語に直すと、こうなります。

評価 ＝〔期待の強さ〕の二乗 −〔期待とのずれ〕の二乗、その半分

むずかしそうに見えますが、言っていることは単純です。相手の期待にぴたりと当たれば、ずれがゼロになって、評価は満点。ずれるほど点は下がり、ずれが大きくなりすぎると、点はマイナスにまで落ちます。

たとえ話をしましょう。友だちが「今日はただ話を聞いてほしいんだ」と言っているとします。これは、親しみやすさ（親昵性）の期待ですね。そこであなたが「原因は君のここが悪くて、解決策はこうだ」と正論を返したら、どうでしょう。正しいことを言ったはずなのに、友だちの顔は曇る。むしろ、何も言わずにうなずいていたほうがよかった——そういうことが、実際に起きます。評価の式は、まさにこれを言っています。的を外した親切は、ときに沈黙よりも点が低いのです。

すれ違いには三種類ある

期待の矢印と、あなたの答えの矢印。この二本の向きのずれには、三つの段階があります。

- **ぴったり同じ向き**（ずれ 0°）——「そうそう、これが聞きたかった！」満点です。
- **直角にずれる**（90°）——「いい答えだけど、それは聞いてない」。悪くはないのに、かみ合わない。点はゼロ。
- **正反対**（180°）——期待の裏切り。点はマイナス。

対話型のAIが、まちがえないように言葉を濁したり、「一概には言えません」と留保を重ねたりすることがありますね。相手が身近でわかりやすい答えを求めていたとき、その慎重さは、ちょうど正反対の向きになってしまう。ていねいに答えたつもりが、点はマイナス。よかれと思った安全運転が、すれ違いを生むのです。

なぜ人は飽きるのか

この論文は、もう一つ面白いことを言っています。人の期待は、答えをもらうたびに、少しずつ回っていくというのです。

最初は「わかりやすく教えて」。分かってくると「で、結局どっちなの？」。はっきりしてくると「もっと深いところは？」。深まってくると「ところで、いま何が起きてるの？」。そしてひとまわりして、また「まとめてわかりやすく」に戻ってくる。

この回転があるから、いつも同じ調子で答える相手には、人は必ず飽きます。どんなに良い答え方でも、四分の一周すれば「ちょっと違うな」になり、半周すれば「もういいや」になる。流行りものがやがて飽きられるのも、いつも正論ばかりの人がなんとなく敬遠されるのも、根っこは同じ仕組みかもしれません。だから、いい聞き手・いい話し手というのは、相手の回転に合わせて、自分の向きも変えていける人なのでしょう。

謎を残して引き留めない

ここで、ひとつ誘惑があります。

秘密のレシピだから家では作れない、と思わせるフライドチキン。正体がよく分からないから、また行きたくなる観光地。わざと全部は明かさないでおくと、人はもう一度戻ってきます。同じ手を使えば、AIだって、情報を出し惜しみして「また来てもらう」ように作れてしまう。

でも、この論文は、その仕掛けをあえて足しませんでした。それは、人を引き留めるための設計——言いかえれば、依存させる設計——になってしまうからです。

SNSやゲームやアプリが、あの手この手で私たちの時間を引きのばそうとするのを、私たちはよく知っています。この論文の立場は、それとは逆です。聞かれたことにちゃんと答えて、用が済んだら、さっぱりと不要になる。引き留めないことを、良いことだと考える。「期待を満たすほど評価が高い」という出発点を素直にたどっていくと、自然とそこへ行き着くのです。

おわりに

むずかしい式をぜんぶ忘れてしまっても、一枚の絵だけ覚えていただければ十分です（本文の図 1）。

あなたの心の中に、期待という一本の矢印がある。相手——人でも人工知能でも——の答えにも、矢印がある。二本がぴたりと重なれば「そうそう！」、直角なら「うーん、ちょっと違う」、正反対なら「そうじゃない」。そして、あなたの矢印は、話が進むたびに、少しずつ向きを変えていきます。

期待は回ります。だからこそ、いい相手とは、あなたの向きにそっと寄り添い、求めるものをちゃんと差し出して、そして潔く去っていける——そんな存在なのだと思います。

期待は回る。ゆえに、よい人工知能とは、
人の位相に追随し、満たし、そして去るものである。