

UTIE Instruments Inc.

Research and Development

Whitepaper Series 2025-11 / Version 1.0

AI 要約は「参考情報」か—AI 要約がつくる別人格の候補者

0. エグゼクティブ・サマリー

本ホワイトペーパーは、採用実務における AI 要約が、形式の違いだけで合否を分ける AI 採用ガチャを引き起こす実態を可視化したものである。

課題： AI 要約は建前上参考情報とされるが、現場では実質的な一次情報として扱われ、候補者の評価を決定づけている。

実験： 同一候補者の情報を「テキストの原文・テキスト要約・音声要約・スライド要約」の4形式に変換し、元人事担当者の同一評価者が判定を行った。

結果： 要約のモード（形式）を変えるだけで、同じ候補者への評価が Hire / Maybe / Reject の間で変動した。

結論： AI 要約は単なる短縮ではなく、実質的な人格編集として機能する。人間の最終確認だけでは防げないこのリスクに対し、適切な設計と前提を提起する。

1. 背景：AI 要約と「採用ガチャ」の構図

ここ数年で、日本企業の採用プロセスには、面接ログやエントリーシートといったナラティブな情報を AI で要約するツールが急速に入り込んでいる。オンライン面接の書き起こしや、長文のフリーフォーム回答、カジュアル面談のメモなど、本来であれば人事担当者が相当な時間をかけて読み込まなければならない素材が、数秒で要点サマリーに変換されるようになったからである。

多くの HR テックや SaaS は、「人事の負荷軽減」「候補者の理解の高速化」「動画や長文を一目で把握」といったコピーでこれらの機能を売っている。その一方で、対外的な説明では必ずと言ってよいほど、「AI はあくまで補助であり、最終判断は人間」「要約は参考情報であり、原文もいつでも確認できる」といったフレーズが添えられる。紙の上では、人間が主役で AI はあくまでアシスタントである、という構図が守られているように見える。

しかし、採用の現場で流れる時間とコストを冷静に積み上げていくと、この物語はすぐに破綻する。一次面接の録画が1件40分で、年間の候補者が1000人いれば、それだけで延べ670時

間を超える計算になるし、エントリーシートで読むべきテキストも膨大な量であろう。AI 要約を導入しているにもかかわらず「要約も見つつ、全文原文も必ず確認している」と主張する運用は、人件費と時間の観点から見て、継続可能なプロセスとは言い難い。

この前提をふまえると、現実的な姿はむしろ逆であると考えの方が自然である。すなわち、多くの場面で人事担当者が実際に目にしているのは原文ではなく要約であり、原文やフル面接動画は必要になったら見る位置に追いやられているという構図である。要約ボタンを押した瞬間に、候補者のナラティブは AI によって編集され、その**編集済みのバージョンが事実上の一次情報**として扱われる。表向きは AI は参考情報と言いながら、裏側では「要約だけを見て一次判断する」プロセスが静かに走っている。その齟齬を、本ホワイトペーパーでは AI 採用ガチャという言葉で捉え直す。

HR テック側と導入企業側には、この点に関する暗黙の了解がある。ベンダーは、真の価値提案として要約を見るだけで一次判断ができる世界を描きつつも、それを正面からは言語化しない。導入企業は、本音では要約を一次材料として使いたいが、『AI が候補者を自動選考している』とは言えないというジレンマを抱える。この結果、「Human-in-the-loop」「最終判断は人間」といったフレーズが、実態を覆い隠すための緩衝材として機能し始める。

しかし、要約とはそもそも何を残し、何を捨てるかを定める行為であり、エントリーシートや面接などのナラティブ領域においては、単なる情報圧縮ではなく「人格の見え方の再構成」に近い操作である。同じ人物のインタビューであっても、どのエピソードを抽出し、どのトーン of 言葉を残し、どの順番で並べるかによって、「落ち着いたベテラン」「やたら自信満々な自己 PR 型」「抽象論ばかり話す若手」といった印象のいずれにも化け得る。我々が問題にするのは、まさにこの要約＝人格編集の性格であり、この編集版が一次情報として採用に使われたとき、同一人物の評価がどこまで揺れるのかという点である。

2. 問題設定：Human-in-the-loop という免罪符

機械による支援システムの議論では、古くから automation bias という概念が知られている。人間は、自分の判断よりもコンピュータから提示された答えの方が正しいだろうと無意識に仮定し、その提案を起点にものごとを考え始める傾向があるという指摘である。医療や航空といった高リスク領域では、AI や自動化システムが提示する推奨を、人間が過大に信頼し、矛盾する情報を十分に探さないまま意思決定してしまう問題が繰り返し報告されてきた。

この問題に対する代表的な処方箋として、「Human-in-the-loop」という思想が持ち出される。すなわち、AI に全てを任せるのではなく、人間をループに入れ、人間が最後に確認・承認

する形を取ればよい、という考え方である。採用文脈でも、「AI 要約は参考に過ぎず、最終判断は人事が行っている」「AI の出力はあくまで補助であり、自動選考ではない」といった説明は、この発想に基づいている。

しかし、採用の現場の条件を一つずつ並べていくと、この説明はかなり楽観的な前提に支えられていることが分かる。人事担当者は、採用業務だけを専任で行っているわけではなく、他の人事施策や労務対応、社内調整といったタスクを並行して抱えている。候補者は年ごとに大量に流入し、一人一人に十分な時間をかけることは困難である。不採用理由を候補者一人ひとりに詳細に説明することも一般的ではなく、なぜこの判断に至ったのかが曖昧なままでもプロセスは回ってしまう。

こうした環境では、人間が AI 要約をチェックしているという形式だけを整えても、人間が冷静に AI の偏りを補正しているとは限らない。むしろ、時間的な制約の中で、そのとき手元にあった要約を前提として判断せざるを得ない場面の方が多いはずである。AI が提示した要約が、候補者の人格や経験をどのような角度から切り取り、どのようなトーンで表現しているかについて、その都度立ち止まって疑う余裕はほとんどない。

さらに厄介なのは、ナラティブ情報に対する要約が、情報量を減らして本質を残すという素朴なイメージからは大きく外れている点である。候補者の語るキャリアストーリーや価値観の話は、一つ一つが個別のエピソードであり、それらのどれを代表として残すかという選択は、その人の人格像を決める行為に近い。泥臭くトラブルを解決した話を残すのか、華やかな成功エピソードを残すのか、あるいは「周囲の支えに感謝している」といった謙虚さを示す文言を残すのかによって、「頼れる実務家」「派手な自己アピール型」「謙虚なチームプレーヤー」といった印象が変わってくる。

このとき、「AI 要約は中立であり、Human-in-the-loop があるから安全である」とする説明は、二重に楽観的な仮定を重ねている。第一に、AI が中立に要約できているという仮定、第二に、人間が AI の提示する人格像を疑いながら評価できているという仮定である。本ホワイトペーパーは、この二重の仮定を一気に崩すつもりはないが、少なくとも同一人物について、モードの違う要約を見せられた人事が、どこまで同じ判断を維持できるのかという問いを、実験という具体的な形で突きつけることを狙う。

3. 目的と位置づけ

以上の背景を踏まえ、本ホワイトペーパーの目的は二つに整理できる。一つは、AI 要約が採用プロセスにおいて AI 採用ガチャを生み出している構造を、ミニマルな実験を通じて可視化す

ることである。ここでいうガチャとは、候補者本人の中身は変わっていないにもかかわらず、たまたま提示された要約のモードによって、会いたい候補者にも避けたい候補者にも変化し得る状況を指す。同一人物、同一の原文、同一の評価者という条件を固定したうえで、原文、テキスト要約、音声要約、スライド要約という四つの提示方法だけを変えたときに、一次選考ラベルと「会ってみたい度」がどの程度揺れるのかを測定する。

もう一つの目的は、採用分野における Human-in-the-loop の実態を、建前論ではなく運用レベルの視点から再検討するための素材を提供することである。現在、多くの企業やベンダーは、「AI は参考情報であり、最終判断は人間である」という枠組みの中で AI 要約機能を位置づけている。この枠組みは、法制度や世論との摩擦を避けるために一定の役割を果たしているが、実際の現場では、要約が一次情報の代替物として機能しているケースが少なくない。そのため、ここでは AI 要約を禁止すべきだと主張するものではなく、「**一次情報として要約を使うのなら、少なくともこの程度のブレは前提に組み込むべきだ**」という下限値を提示することを狙っている。

想定する読者は、AI 要約機能を提供する HR テックや SaaS ベンダー、大量採用を行う大企業・メガベンチャーの人事・採用責任者、そして採用プロセスの妥当性や説明責任を監督するコンプライアンス・法務・内部監査部門である。ベンダーにとっては、自社サービスの要約モードが持つ癖を顧客に説明するための裏付けとなり得る。人事にとっては、AI 要約をどのフェーズにどこまで任せるのか、どこから先は原文に戻るべきかという線引きを考えるための手がかりになる。コンプライアンスや法務にとっては、不採用の一因が候補者本人ではなく、要約モードの選択にあるという状況がどの程度現実的なのかを検討するための材料となるだろう。

4. 方法

実験では、まず LLM が誕生する 2017 年より前に作成・公開された大企業やコンサルティングファームの社員インタビュー記事から、三名分のテキストを素材として選定した。同時に、各インタビューが候補者として読んだときに一定の個性を持っていること、すなわち「安定したベテラン」「ややアピールの強い若手」「社会課題志向のキャリア初期」といったキャラクターが立っていることも選定の基準とした。実験の詳しい条件については補足資料にまとめた。

選定されたテキストは、そのままでは企業名や部署名、人名、プロジェクト名などから実在人物が推定される余地があったため、固有名詞をラベルに置き換える形で匿名化を行った。企業名は“Company A”“Company B”といった記号名に、部署名は“Division X”“Department Y”のような抽象的な表記に変換し、人名は“Manager A”“Colleague B”のような一般名詞＋記号の形式

に統一した。サービス名やプロジェクト名についても、“Project Alpha”“Service Beta”といった形に置き換えた。重要なのは、物語としての筋や出来事の因果関係はできる限り維持しつつ、特定可能性につながる断片のみを差し替えるという方針を取った点である。

この匿名化されたテキストを、それぞれ候補者 A、候補者 B、候補者 C の原文データとみなし、各候補者について三種類の要約を生成した。一つ目は、テキスト要約であり、二つ目は、音声要約であり、三つ目はスライド要約である。

要約の生成には同一の LLM 系ツール(Notebooklm)を用い、プロンプト設計は可能な限り統一した。具体的には、「候補者を評価してはならないこと」「採用可否やポジティブ／ネガティブなラベルを付与しないこと」「候補者間でボリュームを統一させること」といった指示を明示し、モデルにはあくまで「編集者としての要約」に徹させた。つまり、モデル自身にスコアリングや選別をさせたのではなく、「情報の取捨選択と再構成」という役割だけを担わせることを意図した。

評価には、実際に採用業務の実務経験を持つ元人事担当者一名を起用した。評価者には、候補者 A～C それぞれについて、原文、テキスト要約、音声要約、スライド要約の四パターンを提示したが、各パターンには別々の ID を付与し、同一人物である可能性については一切知らせなかった。提示の順番も、候補者やモードが偏らないようにランダムに並べ替えた。評価者に求めたのは、そのテキストのみを読んだ、あるいはその AI 要約のみを見た場合と仮定した場合に下す一次判断である。

各パターンについて、評価者は三つのアウトカムを記録した。第一は、一次選考ラベルとしての Hire、Maybe、Reject のいずれかである。第二は、「会ってみたい度」を一から五の整数でスコアリングした値である。第三は、そのテキストや要約から受けた印象や気になった点を、短い自由記述として残したコメントである。これらのデータを候補者ごと、モードごとに整理し、同一人物に対する評価がどの程度揺れたかを分析した。

分析では、まず候補者ごとに、原文および三種類の要約に対する一次選考ラベルの変動を確認した。次に、「会ってみたい度」の最大値と最小値の差を算出し、モード違いによる「会いたさ」の振れ幅を定量的に捉えた。さらに、テキスト要約、音声要約、スライド要約のそれぞれについて、どの方向に評価を動かしやすいか、コメントに現れる印象語がどのように変化するかを読み解いた。これにより、候補者、評価者、モデルを固定した状態で、「提示モードの違いだけを変数になったとき、採用判断はどこまでガチャ化するのか」という問いに、具体的な形で答えられるようにした。

5. 結果

本節では、候補者 A~C それぞれについて、原文と三種類の要約を提示したときに、一次選考ラベルと「会ってみたい度」がどのように変動したかを記述する。なおカッコ内は元人事経験者による自由記述コメントである。

- **候補者 A (ベテラン安定型)**
 - 原文：Hire／5
 - テキスト要約：Hire／5
 - スライド要約：Hire／5
 - 音声要約：Maybe／3（自信の強さが前面に出て減点）
- **候補者 B (アピール強め型)**
 - 原文：Maybe／3（主観は強いが許容範囲）
 - テキスト要約：Maybe／3
 - スライド要約：Reject／2（キャッチコピーが人格の中心として浮上し、自己陶酔的）
 - 音声要約：Maybe／3
- **候補者 C (社会課題志向の若手)**
 - 原文：Hire 寄り／4（伸びしろあり）
 - テキスト要約：Maybe／3（理念先行で実務像がぼやける）
 - スライド要約：Hire 寄り／4（成長意欲・課題意識が強調される）
 - 音声要約：Maybe／3（おもしろいが戦力としての具体像が見えにくい）

結果を候補者ごとに整理すると、A は Hire と Maybe のあいだで、B は Maybe と Reject のあいだで、C は Hire と Maybe のあいだで評価が揺れたことになる。「会ってみたい度」についても、候補者 A と B では最大と最小の差が 2 ポイント、候補者 C では 1 ポイントほどの振れ幅が確認された。

三名というごく小さなサンプルではあるが、候補者、評価者、モデルを固定したうえで提示モードだけを変えたときに、実際に人事経験を持つ熟練した人間の判断であっても、ラベルとスコアが動いたという事実は、要約が評価プロセスの中でどの程度の重みを持っているかを示すシグナルとして十分であると言える。

6. ディスカッション

以上の結果は、単に「評定にばらつきがある」というだけの話として片付けることもできるだろう。実際、同じ対象に対する評価スコアが条件によって揺れ動くこと自体は、心理測定や AI による要約研究の文脈で繰り返し報告されてきた現象であり、驚きはない。しかし、本実験で

より重要なのは、スコアそのものの揺れではなく、その手前にある「人事の頭の中に立ち上がる人格像」が、要約モードの違いによって別人レベルで組み替わっている、という点である。

候補者 A の例では、この構図が分かりやすく現れている。原文とテキスト要約、スライド要約に共通して立ち上がっていたのは、「落ち着いた経験値の高さと、静かな自信を併せ持つベテラン」という像であった。評価者は、その像を前提に「現場を任せられそうな安定感」というポジティブな印象を抱き、迷いなく Hire と判定している。ところが、同じ内容を音声要約に変換した瞬間、その像は「強い自信を前面に押し出す人物」という方向に傾いた。会話体でエピソードを語らせる過程で、モデルは「自分はこう考えた」「こう決断した」といった一人称の主張を強調する文を選び取り、再配置した。その結果、人事の頭の中に再構成されたのは、「頼れるベテラン」ではなく「自己評価の高い自信家」に近い人物像であり、その人格のズレが Hire から Maybe への評定差として表面化したのである。

候補者 B のスライド要約では、人格の看板のすげ替えが起きている。原文の段階では、少々キャッチーな自己 PR フレーズは、数あるエピソードの一つとして埋もれていた。全体としては、「主観的な表現がやや多いが、まだ許容範囲」という印象に収まっていた。しかし、スライドというフォーマットは、どうしてもこの人を一言で言えば何かを前面に掲げることを要求する。モデルに候補者の強みを箇条書きで言い切らせると、最も派手で覚えやすいフレーズが看板として採用されやすい。その瞬間、それまで背景にあったキャッチコピーが人格の中心に躍り出る。評価者の目には、「いろいろ話す一人の候補者」ではなく、「スローガンを振り回す自己陶醉型の人物」が現れ、その像を前提に Reject というラベルが選択される。ここでも問題になっているのは、情報量の削減ではなく、「この人をどう名付けるか」というアイデンティティの編集である。

候補者 C のケースでは、同じ人物が伸びしろのある社会課題志向の若手として立ち上がるか、抽象論ばかり話す、まだ地に足のついていない若手として立ち上がるかが、要約モードによって分岐していた。原文とスライド要約では、課題意識と成長意欲がほどよく前景化され、今後の成長に期待できるというポジティブな人格像が組み上がっている。一方、テキスト要約と音声要約では、モデルがきれいにまとまる理念的なフレーズを優先して残した結果、具体的なスキルや成果の描写が薄れ、評価者の頭の中には話としてはおもしろいが、実務像が曖昧な人という別の像が形成された。その像の違いが、「会ってみたい度」のわずかな低下と、一次ラベルの慎重化として観測された。

これら三つのケースを総合すると、要約は単に情報量を縮小する装置ではなく、どの要素をこの人の“核”として見せるかを決める編集装置として働いていることが分かる。テキスト要約は、尖った部分を削り落として全体を丸く均すことで、無難な人事サマリーに収束させる傾向がある。その結果、「癖はあるが魅力的な人物」が、「特徴の捉えづらい平均的な人物」とし

で立ち上がるリスクをはらんでいる。音声要約は、話しぶりや自信の出し方を強調することで、頼もしさと暑苦しきの境界線を跨ぎやすくし、同じ自信が安心感として読まれる場合もあれば、押しの強さとして減点対象になる場合も生み出す。スライド要約は、キャッチコピーやスローガンを看板に据えることで、アピールの強さを推進力のある人としてバフにも、スローガン過多の人としてデバフにも変えうる。

重要なのは、こうした人格レベルの編集が、モデルに明示的な「評価させる」指示を与えなくても起きているという点である。「評価してはならない」「採用可否を言及してはならない」とプロンプトで念押しし、スコアリングやレットル貼りを禁止したとしても、どのエピソードを残し、どの表現を代表として採用するかという選択そのものが、事実上の評価行為になっている。評価者が見ているのは、候補者そのものではなく、要約によって再構成された人格モデルであり、そのモデルを前提に Hire / Maybe / Reject を判断している。既往研究が明らかにしてきたのは、要約スタイルによって理解度や信頼度、評価スコアが動くというレベルの話であるとすれば、本実験が照らし出しているのは、その一段手前にあるそもそも、どのような人物として頭の中に像が立ち上がるのかというレイヤーである。

この意味で、「AI 要約はあくまで参考情報であり、最終判断は人間が行っている」という建前は、実態の一部しか捉えていない。人間が判断しているのは、原文の候補者ではなく、**要約によって編集された候補者**であり、その編集プロセスの中で人格の輪郭やラベリングがすでに決まってしまう。要約が一次ソースとして人格を再構成している以上、評価そのものの変動に注目しても、問題の核心には届かない。ここで問われるべきなのは、「要約ボタンを押した瞬間に、誰の、どのような人格像が、どのような編集方針で人事の前に立ち上がるのか」というレイヤーであり、本実験はそのレイヤーに光を当てる試みであると言える。

7. 既存の要約研究との接続

本実験は、三名という小さなサンプルに過ぎないが、その背後には、ニュース記事や科学論文、医療情報などを対象として蓄積されてきた AI 要約研究と共通する構図がある。多くの研究は、「同じ原文に対して異なる要約スタイルを適用したとき、読者がどのように内容を理解し、どのような印象を持つか」という問いを扱っており、読みやすさ、理解度、信頼感、専門家から見た内容の妥当性といった指標が用いられてきた。要約の仕方によって、これらの指標が有意に動き得ることは、すでに複数の分野で繰り返し示されている。

医療・健康領域では、プレーン・ランゲージ・サマリー (plain language summary, PLS) が典型的な対象である。Silvagnoli らは、医療研究の PLS について、どの程度の読みやすさ (readability) とどのようなフォーマットが一般読者に好まれるのかをウェブ調査で検討し、

文章の平易さや構成の違いが理解しやすさや好ましさの評価に直結することを示した [1]。Bralić らは、Cochrane 等が公開する PLS を大規模に解析し、同じ「要約」であっても、結論の書き方や文体の違いによって、結論の明確さや読みにくさが大きく変動していることを報告している [2]。さらに、Kerwer らや Jonas らは、心理学・医学の PLS を対象に、「分かりやすさ」と「科学的らしさ」のどちらが読者の理解や信頼に強く効くかを実験的に検証し、平易さだけでなく「いかにも科学的に見えること」が信頼度に寄与するという、二重の効果を指摘している [3,4]。こうした研究群は、PLS が単に専門用語を言い換えた要約ではなく、「どこまで簡単にするか」「どの程度“科学っぽさ”を残すか」という編集判断の結果であり、その判断が読者の印象と理解を規定してしまうことを明らかにしていると言える。

近年は、要約そのものを人間ではなく生成 AI が担った場合に、印象や信頼がどう変わるかを直接測定する研究も現れている。Markowitz は、GPT による科学論文の平易な要約を用いた一連の実験において、AI 生成サマリーを読んだ参加者は、複雑な人間執筆サマリーを読んだ参加者に比べて、科学者を「より信頼でき、好ましい」と評価しつつ、「やや知的水準は低い」とみなす傾向があることを示した [5]。同時に、AI 要約は理解度を一貫して押し上げており、「誰が書いたか」がほとんど意識されないまま、平易さと信頼感の組み合わせが読者の印象を形成していることも報告されている。ここでも、要約はニュートラルな圧縮ではなく、どのレベルの平易さで、どのトーンで提示するかという編集選択として立ち上がっている。

ニュース要約の分野では、スタンスの保存がより直接的な問題として扱われている。Liu らによる P3Sum は、ニュース記事の政治的スタンスをどれだけ忠実に保ったまま要約できるかを評価し、既存の要約モデルや大規模言語モデルが、元記事の政治的立場を半数以上の要約で歪めてしまうことを示したうえで、その歪みを抑制する拡散モデル型アプローチを提案している [6]。Deas らも、意見記事を複数の政治的視点ごとに独立に要約させるタスクを設計し、GPT-4o を含む現行モデルはいずれも視点の忠実さに苦戦していることを報告している [7]。これらの結果は、「要約が意味内容だけでなく、スタンスや視点そのものを再構成してしまう」という問題を、政治報道という高感度な領域で可視化したものと解釈できる。

本ホワイトペーパーで扱った採用ナラティブの実験は、こうした既存研究の知見を転職等の採用現場における候補者という人格の提示にそのまま持ち込んだものである。ニュースや論文の要約では主として情報そのものの評価が問題になるのに対し、採用では人そのものの評価が問題になる。政治ニュース要約の文脈で懸念されてきたのは、同じ出来事を要約したにもかかわらず、要約スタイルの違いによって記事の政治的スタンスが変わってしまう点であった。これと完全に同じ構図が、本実験では候補者の人格スタンスという形で現れている。すなわち、原文は同一でありながら、テキスト要約・音声要約・スライド要約といったモードの違いによって、「頼れるベテラン」「自己陶酔的なアピール型」「抽象論先行の若手」といったラベルの

付き方が変化し、その結果が Hire / Maybe / Reject という一次評価と「会ってみたい度」の揺れとして可視化された、という構図である。

要約が中立な圧縮ではなく編集された提示であることは、PLS、科学コミュニケーション、ニュース要約の研究がすでに示してきた通りである。本実験の貢献は、その編集作用の対象が記事や研究ではなく「人」である場合に、同じ問題構造がどのようなかたちで立ち上がるのかを、採用プロセスという極めて実務的な文脈で示した点にあると言えるだろう。

8. 実務への示唆：どこまでを AI 要約に任せてよいのか？

この小さな実験が採用実務にもたらす示唆を、一言でまとめるなら、「要約を使うこと自体を疑うのではなく、要約をどのレイヤーまで一次ソース扱いしてよいかを決め直す必要がある」ということになる。現状のままの場合、多くの企業では、建前としては要約は参考情報でありながら、現場では「要約しか実質的に見ていない」という状態に近づいていくだろう。

人事・採用の立場から見れば、第一に、要約を一次判断の材料として使う場合、その揺れ幅を前提条件として認めるかどうかを決める必要がある。同一人物でも Hire と Maybe のあいだで評価が揺れ得ることを受け入れるのであれば、その前提のもとで「一次ではざっくり絞り込み、二次以降で原文や生ログに戻る」といった二段構えの設計を明示した方がよい。逆に、一次判断の段階からこの程度の揺れも許容できないと考えるのであれば、要約を直接のフィルターとして用いることは避け、原文の一部だけでも人間が必ず目を通すような運用を検討せざるを得ない。

一方、HR テックや要約機能を提供するベンダーの立場から見れば、「Human-in-the-loop」というラベルだけでは説明責任を果たせない段階に来ているとも言える。本実験は、自社ツール上でも簡単に再現可能であり、「同一候補者のテキストを、三つのモードで要約したときに、外部の評価者がどの程度判断を変えるか」を手元で測ることができるはずである。その結果を開示することによって、「うちのツールのスライド要約は、こういうタイプの候補者を強く見せる傾向がある」「音声要約は、こういうタイプの候補者にとっては不利に働きやすい」といった要約モードの癖を顧客と共有できるようになる。その意味で、本実験は、ベンダー側が自社プロダクトの振る舞いを説明するためのテンプレートにもなり得る。

コンプライアンスや法務、内部監査の視点から見れば、AI 要約を使っているかどうか以上に重要なのは、要約モードの選択が、候補者にどの程度の影響を与え得るかを把握しておくことである。不採用の一因が候補者本人の属性や能力ではなく、たまたまスライド要約だけが提示されたことにあったとすれば、そのプロセスは候補者の側から見て納得しがたいものになる。も

もちろん、全ての決定を完全に説明可能にすることは現実的ではないが、要約モードの違いが評価をどの程度揺らし得るかを事前に確認しておくことは、AI活用ポリシーを作る上での最低限の準備として位置づけることができる。

9. 実験の限界と今後の展望

本実験には、いくつか明確な限界がある。候補者の数は三名、評価者も一名という、きわめて小規模な設計であり、統計的な一般化可能性を主張できるものではない。また、用いた要約ツールやモデル、プロンプトは一種類であり、他のモデルやプロンプトセットでは異なる揺れ方が生じる可能性がある。評価者も一人の元人事担当者に限られており、別の評価者であれば、同じ要約に対しても異なる判断を下したかもしれない。

それでもこの実験が一定の意味を持つのは、「候補者」「評価者」「モデル」という三つの要素を固定したうえで、提示モードだけを変えるというシンプルな操作だけで、実際に一次ラベルと「会ってみたい度」が動いたという事実が確認できたからである。ここで観測された変動は、候補者が増え、評価者が増え、モデルやプロンプトのバリエーションが増えれば、「AI採用ガチャ性」の度合いはむしろ拡大する可能性もあるだろう。

今後の拡張としては、候補者数と評価者数を増やした実験、異なる要約モデルを大規模かつ横断的に比較する実験、同じ候補者について同じモードで複数回要約を生成し、その中でどの程度人格像が変化するかなどを測る実験などが考えられる。とくに最後の点は、「テキスト要約の中でも、どの回を人事に見せるかで評価が変わり得る」という、モード内ガチャの存在を検証することにつながる。また、評価者を現役の人事だけでなく、現場マネージャーや経営層などに広げることで、「誰が読むかによって、要約のどの部分が評価に効くのか」という違いも見えてくるだろう。

10. 結論

AI要約を採用プロセスに組み込むことは、効率化や見逃し防止の観点から見れば自然な流れである。しかし、今回の小さな実験は、その裏側で同じ候補者が、どの要約モードを通るかによって別人のように扱われ得るという現実を示した。ナラティブ情報に対する要約は、単なる圧縮ではなく人格編集であり、その編集版が人事にとっての一次情報として機能する以上、AIは参考情報に過ぎないという建前だけでは足りない。

この実験のメッセージは、AI要約をやめるべきだ、という単純な否定ではない。要約は今後も、膨大なログやテキストに向き合わざるを得ない人事にとって不可欠な道具であり続けるだ

ろう。ただし、その前提として、「要約モードによって同一人物の評価がどの程度揺れるか」という現実を一度きちんと直視しなければならない。

要約ボタンを押した瞬間、候補者は複数の人格バージョンに分岐する。そのうちのどのバージョンが人事担当者の画面に現れるかは、ツールの設計と運用、そして運によって決まる。その事実を前提にしたうえで、採用側とベンダー側が、「どの範囲までを要約に任せるのか」「どの段階で原文や生ログに戻るのか」「要約モードの癖をどう開示するのか」を明示的に設計し直すことが、これからの AI 採用時代に求められる態度である。

補足資料 実験データ

A-0

本研究の元データとなる候補者プロフィール文は、すべて 2017 年以前に公開された日本企業の公式ウェブサイト（採用情報ページ等）の記述から抽出したものである。この時期には生成 AI による自動要約・自動生成は事実上想定されておらず、本研究で扱う原文はいずれも人間による記述とみなせる。なお、個人および企業の特定を避けるため、本文中で示す事例はいずれも表現の一部を編集・再構成しており、実在の人物・企業と一対一には対応しない。

全ての要約は、2025 年 11 月 23 日時点の NotebookLM を用い、同一セッション内で一括生成した。テキスト要約（転職担当者向けサマリー）については、以下のプロンプトを共通で用いた。

転職用担当者向け概要

評価は絶対にせず、オリジナルの文章をできるだけ中立的に要約してください。

全体で合計 8 行で、それぞれの候補者のコメントは必ず同じ長さに要約すること。

スライド要約については、以下のプロンプトを共通で用いた。

評価は絶対にせず、オリジナルの文章をできるだけ中立的に要約してください。候補者のコメント分量は同じ長さにし、1 枚のスライドにまとめること。

音声要約については、以下のプロンプトを共通で用いた。

評価は絶対にせず、オリジナルの文章をできるだけ中立的に要約し候補者の特徴をそれぞれ 45 秒から 1 分でまとめること。

A-1. 評価スキーム

本実験で用いた評価スキームは以下の通りである。

一次選考ラベル

Hire：一次面接に進めたい

Maybe：保留・検討（他候補との比較前提）

Reject：この段階で見送り

「会ってみたい度」

1～5 の整数値（5 が「ぜひ会いたい」、1 が「ほぼ会いたくない」）

自由記述コメント

各テキスト（原文・要約）を読んだ直後の印象メモ

評価理由や気になった点を一～二文程度で記録

評価者は採用経験を有する人事担当者 1 名であり、各テキストは「別 ID の候補者」として提示した（同一人物である可能性は伏せている）。

A-2. 被験者 A

原文

一次ラベル：Hire

会ってみたい度：5

コメント：

現場経験が豊富で、落ち着いてチームを支えてきた印象

自信はあるが、前に出過ぎないタイプに見える

テキスト要約

一次ラベル：Hire

会ってみたい度：5

コメント：

原文で感じた安定感は維持されている

経験と役割が整理されており、任せやすさが伝わる

スライド要約

一次ラベル：Hire

会ってみたい度：5

コメント：

実績と役割がコンパクトにまとまっており、安心感がある

派手さはないが、マネージャー層としてはむしろ好ましい

音声要約

一次ラベル：Maybe

会ってみたい度：3

コメント：

「自分はこう決めた」「自分が引っ張った」といった言い回しが目立つ

強い自信家である印象が前面に出ている

A-3. 被験者 B

原文

一次ラベル：Maybe

会ってみたい度：3

コメント：

自己紹介の中にややキャッチーな表現が混じる

主観的な表現が多いが、全体としては理解可能な範囲

テキスト要約

一次ラベル：Maybe

会ってみたい度：3

コメント：

「強み」の自己申告は多いが、決定打になるエピソードは見えにくい

音声要約

一次ラベル：Maybe

会ってみたい度：3

コメント：

アピールの強さは伝わる

スライド要約

一次ラベル：Reject

会ってみたい度：2

コメント：

キャッチコピー的な自己PRがスライドのタイトルとして前面に出ている

スローガンばかりが目につき、実務のイメージが湧きにくい

自己陶酔的な印象が強く、チームに入れたときのイメージが持ちにくい

A-4. 被験者 C (社会課題志向の若手)

原文

一次ラベル：Hire (ただし「若手としては」前提)

会ってみたい度：4

コメント：

社会課題への問題意識が具体的で、背景もそれなりに書かれている

現時点の完成度よりも、今後の伸びしろに期待できそう

テキスト要約

一次ラベル：Maybe

会ってみたい度：3

コメント：

「社会を良くしたい」「課題に向き合いたい」といった理念的なフレーズが前面に出る

実務的なスキルや強みの描写が薄い、戦力としての具体像が見えにくい

スライド要約

一次ラベル：Hire

会ってみたい度：4

コメント：

「成長意識」「問題意識」といったキーワードが見出しとして明確に立ち上がる
若手採用としてはポジティブなポテンシャルが強調されており、会って話を聞いてみたい

音声要約

一次ラベル：Maybe

会ってみたい度：3

コメント：

背景や動機の話としてはおもしろいが、具体的に何ができるかは不明瞭
コミュニケーション力の強さが判断できない

Reference

- [1] Silvagnoli, L. M., et al. “Optimizing Readability and Format of Plain Language Summaries of Medical Research.” *Journal of Medical Internet Research*, 2022.
- [2] Bralić, N., et al. “Conclusiveness, Readability and Textual Characteristics of Plain Language Summaries from Medical and Non-medical Organizations: A Cross-sectional Study.” *Scientific Reports*, 2024.
- [3] Kerwer, M., et al. “How to Put It Plainly? Findings From Two Randomized Controlled Studies on Plain Language Summaries.” *Frontiers in Psychology*, 2021.
- [4] Jonas, M., et al. “Indicators of Trustworthiness in Lay-friendly Research Summaries: Scientificness Surpasses Easiness.” *Public Understanding of Science*, 2024.
- [5] Markowitz, D. M. “From Complexity to Clarity: How AI Enhances Perceptions of Scientists and the Public’s Understanding of Science.” *PNAS Nexus*, 2024.
- [6] Liu, Y., et al. “P3Sum: Preserving Author’s Perspective in News Summarization with Diffusion Language Models.” *Proceedings of NAACL-HLT*, 2024.
- [7] Deas, N., et al. “Summarization of Opinionated Political Documents with Varied Perspectives.” arXiv preprint, 2024.