

支援事業評価の評価効率化と客観化にむけた LLM(対話型 AI)の活用事例

○佐藤浩哉, 小林健輔, 松本匡史, おいとま (岡山大学 研究・イノベーション共創機構)

1. はじめに

拠点形成支援、研究支援、表彰制度などの運営審査への LLM の活用可能性を検討した。「申請書」に対して単に公募要領記載の観点をプロンプトとして適用した“審査”では、審査の意図が必ずしも評点に明確に反映されず、再現性がない。LLM の限界を回避するような運用にあわせ、プロンプトの記述を明確化することで、審査基準に応じた比較・指摘が可能になる。プロンプトの明確化は、各公募等の審査基準の明確化を意味するので、これを制度設計へフィードバックすることで、制度自体の目的の共通認識化と納得性向上につながる。

2. 実在の公募審査の結果と、LLM での評価結果の比較結果

《LLM のシステム条件》 オンプレミスの LLM(Gemma 3 27B¹)

- ・システムプロンプトに評価の観点を読み込めること
LLM のキャラ付けや役割定義と同等の重みで評価の観点を再現性良く徹底するため、採点対象データの前に読み込ませられるようにする。
- ・Temperature=0 設定が可能、かつ、トークンカウントをモニタ可能であること
《適用実験を行った公募案件：大学と外部機関との協力研究への補助の公募》
- ・採択予定 4 件に対し、応募 9 件(うち 1 件は申請後棄権取下げ)
《LLM による評価手順》
- ・個別に採点対象データファイルを準備(合体しない)
- ・システムプロンプトに、評価の観点を含め、すべてのプロンプトを入力。
- ・履歴/順列の影響を避け、またトークン枯渇を防ぐため、毎回リセットした状態から応募書類を投入。ユーザープロンプトボックスには何も記載しない。

《結果 1》プロンプト作成者の判定結果と LLM 評価結果の整合数(9 件中の上位/下位で判定)

簡易版：上位 3 件、下位 3 件

評価ポイント記載版：上位 3 件、下位 3 件

記載強化版：上位 4 件、下位 4 件

⇒プロンプト記載の強化で、作成者の意図をより反映した判断に近づけることは可能

《結果 2》LLM の評価結果で“勝抜け”と“足切り”が実現できる可能性の検証

・目標：LLM 採点結果を基に上位・下位から無条件に採否決定した場合、合議裁定に基づく採否結果と、結果の逆転が発生しない最大件数

簡易版：“勝抜け”1 件 “足切り”1 件

評価ポイント記載版：“勝抜け”3 件 “足切り”2 件

記載強化版：“勝抜け”2 件 “足切り”1 件

⇒9 件中 4 件採択の場合、ポイント記載版なら“勝抜け&足切り”各 2 件程度可能

《合議裁定と個人(プロンプト作成者判定)の評価不整合に関する考察》

合議裁定で当初基準にない視点(社会科学系案件を重視)が出て、#7 が#1 を「逆転」して採択。

#	簡易		ポイント記載		記載強化		プロンプト 作成者判定	合議裁定
	棄権	43	棄権	39	棄権	40		
5	棄権	43	棄権	39	棄権	40	○	棄権
2	○	37	○	39	○	36	○	○
1	○	39	△	36	○	36	○	X
6	○	40	○	37	○	36	○	○
8	X	36	○	39	○	35	△	○
4	X	33	X	32	X	33	X	X
7	X	35	X	35	X	31	X	○
3	X	36	X	32	X	29	X	X
9	○	37	△	36	X	33	X	X

Legend

- AI上位で裁定と整合(勝抜け)
- AIと裁定の不整合
- △ ボーダーで同点
- 合議裁定による採否逆転
- X AIと裁定の不整合
- X AI下位で裁定と整合(足切り)

3. 結論

- ・公募要領/外部公開審査基準に明示された基準に沿った“勝抜け”と“足切り”が実現できる可能性がでてきた。
- ・しかしながら、人の合議は必須。プロンプト作成者のバイアスの修正が必要。また実際の審査員審査では、公募要領に現れない合議ならではの観点が提示される
⇒AI では判定できない LLM の明らかな限界

参考. 具体的なプロンプト例

《システムプロンプト前半部》同一事業の評価中に変更せず

- ・公募要領に記載されている審査観点毎に点数をつけさせ、公募条件に応じて加重評価。
- 【プロンプト例】あなたは、応募者から中立で、誠実で優秀で専門的知識をもつ、厳格な「xx 公募(事業名)」提案書の審査員です。常に日本語で回答してください。回答の各列のヘッダーは下記。
- | 提案番号 | 1)外部資金 | 2)研究分担 | 3)強化手法適切性 | 4)技術的波及効果 | 5)解決方法 | 6)連携先企業等 | 評価理由 |
- 各申請書を行とし、引き続いて投入する pdf ファイルの内容に対して、下記のようなそれぞれの観点で、1) は 20 段階(1~20)評価を、2) から 6) まではそれぞれに対して 10 段階(1~10)評価の素点をつけて。各行に対応して、点数が低く判定された理由を、評価理由欄に 30 - 100 字程度で記載して。

《システムプロンプト後半部の考え方：評価の観点》

【簡易版】

- ・公募要領に記載されている審査の観点を(ほぼ)そのまま記載。
- 例：1) 外部資金 (企業からの資金を含む) 獲得に向けた計画が明確か(獲得目標の明確さ、時期、獲得資金量それぞれについて評価)

【評価ポイント記載版】

- ・審査基準に記載の観点 and/or 提案書に注釈記載された記載指示項目を援用。
- 例：1) 外部資金 (企業からの資金を含む) 獲得に向けた計画が明確か(獲得目標の明確さ、時期、獲得資金量それぞれについて評価)
- 評価ポイント：本公募は、外部資金を取れる(すなわち何らかの形で社会的効用を第三者に認めさせられる)直前の研究を支援することを目的にしている。したがって、外部資金獲得までの課題が的確に把握され、提案された計画(実施内容、スケジュール、実施体制等)が具体的かつ実行可能であることが必要。

【観点記載強化版】

- ・プロンプト作成者の意図を可能な限り明文化することを試みる。
- 例：1) 外部資金 (企業からの資金を含む) 獲得に向けた計画が明確か
- 評価ポイント：本公募は、外部資金を取れる(すなわち何らかの形で社会的効用を第三者に認めさせられる)直前の研究を支援することを目的にしている。したがって、
- ・外部資金獲得目標は明確に設定されているか？複数の資金獲得源を検討していれば加点評価。
 - ・外部資金獲得の目標時期はいつか？2027 年以降なら 1 年あたり 1 点以上減点。2025 年度中なら 1 点加点。
 - ・外部資金獲得目標の金額は明示されているか？明示されていて 30 百万円を超えれば大きく加点評価。5 百万円以下と見積もられるようであれば減点。基盤 C のみでは減点。
 - ・外部資金を獲得するために必要な課題の指摘が明確か？
 - ・指摘されている課題がなぜ資金獲得上の障害になっているかの根拠は明示されているか？
 - ・提案された計画の実施内容は明示的、具体的かつ実行可能か？実施の体制は明確か？
 - ・計画のスケジュールは明示されていて、具体的かつ実行可能か？計画スケジュールに矛盾はないか？
 - ・研究推進の核となる技術について、当該グループの水準を世界水準と比較した結果の記載がなければ減点。高ければ加点。低ければ大きな減点。
 - ・研究の核となる技術を明示し、その技術を保持していることが明示されているか？明示されていない場合は減点。核となる技術の希少性が示されていれば加点。
 - ・保持している技術水準の高さを証明するエビデンスの記載があれば加点。

《参考文献》

1) Gemma 3 Technical Report : <https://arxiv.org/abs/2503.19786>

連絡先 : Sato.hiroya@okayama-u.ac.jp

