

大規模言語モデルによる選好規則の推定と アイデア生成の統合フレームワーク

岩井 皓暉¹ 熊谷 雄介¹ 馬場 雪乃²

¹ 株式会社博報堂 DY ホールディングス ² 東京大学

{koki.iwai,yusuke.kumagae}@hakuhodo.co.jp yukino-baba@g.ecc.u-tokyo.ac.jp

概要

大規模言語モデル (LLM) は、創造的タスクにおけるアイデア生成支援に活用されている。しかし、LLM 単体で生成するアイデアは無難な表現に偏りやすく、個人や文脈に応じた選好を十分に反映できない。また、個人や文脈が変化の中で再学習なしで選好を反映する手法が求められるが、選好基準を指示文として明示することは難しい。本研究では、ペアワイズ比較データを基に LLM が自然言語で選好規則を推定し、その結果を生成プロセスに直接接続する統合フレームワークを提案する。実験では、推定した選好規則により選好の分類精度が向上すること、選好規則に基づいて生成したアイデアは規則なしの場合よりも選好に適合することを示した。

1 はじめに

大規模言語モデル (LLM) は、キャッチコピー制作、商品ネーミング、課題解決アイデア創出など、創造的タスクにおけるアイデア生成支援に活用されている [1, 2, 3]。

しかし、LLM が単体で生成するアイデアは一般的で無難な表現に偏りやすく、個人の価値観や文脈に応じた好ましさ、すなわち (広義の) 選好を十分に反映できていない懸念がある [4]。より実用的なアイデア生成支援のためには、選好を生成プロセスに反映させることが重要である。

一方で、アイデア生成支援の利用場面では対象となる個人や文脈が頻繁に変化するため、その都度モデルを再学習・調整する手法は運用上現実的ではない。このため、事後学習を伴わず、指示文の調整により選好を柔軟かつ軽量に反映できる手法が求められる。ただし、指示文によって選好を反映しようとした場合、個人が自身の選好の基準を明示的に言語化することは容易ではない。特に創造的なアウト

プットを対象とする場面では、「なぜそのアイデアを好むのか」といった選好理由の言語化が難しく、指示文による直接的な選好指定には限界がある [5]。

本研究では、選好を直接言語化させるのではなく、ペアワイズ比較 (2つのアイデアのうちどちらが好ましいかの判断) という比較的容易な判断から、LLM が選好の基準を自然言語で表した選好規則を推定し、その結果を生成プロセスに直接接続する統合フレームワークを提案する。選好規則を自然言語として明示的に表現することで、個人が暗黙的に持つ選好を可読かつ再利用可能な形で扱うことができ、推定された規則をそのまま生成条件として LLM に与えることが可能となる。

実験では、推定した選好規則を用いることで選好の分類精度が向上すること、および選好規則に基づいて生成したアイデアが、規則を用いない場合と比べてより選好に適合することを示す。

2 関連研究

LLM による選好を反映した生成 (Personalized Generation) に関する研究は活発に行われており、体系的なサーベイも報告されている [6]。本研究は、選好の解釈可能な表現、選好へのアラインメント、および条件付けによる生成制御という3つの研究潮流にまたがる立場から、関連研究を整理する。

2.1 選好の解釈可能な表現

人間の選好を、解釈可能な構造や表現として捉えるための手法が提案されている [7, 8]。特に、Findeisらは、ペアワイズ比較データから選好の背後にある原則を自然言語で抽出する手法を提案し、選好データの解釈可能性を示した [8]。しかし、これらの研究は選好の表現や解釈を主目的としており、推定した選好を生成タスクの制御に直接用いる枠組みは検討していない。

2.2 選好へのアラインメント

人間の判断データに基づきモデル出力を選好にアラインメントする学習ベース手法として、Reinforcement Learning from Human Feedback (RLHF) [9, 10] や Direct Preference Optimization (DPO) [11] が提案されている。また、個人差を扱う拡張として、潜在変数を用いた Personalized RLHF も提案されている [12]。これらは追加学習を前提とする点で、本研究が対象とする指示文ベースの軽量の枠組みとは設計思想が異なる。

2.3 条件付けによる生成制御

生成条件によって出力を制御する手法として、事前定義した制御コードを用いる CTRL [13] や、外部分類器のスコアにより生成を誘導する PPLM [14] が提案されている。また、推論時に自然言語の指示や例を与える文脈内学習 [15] に基づく生成制御も提案されている [10, 16]。一方、これらの手法はいずれも、人手で設計された制御コードや指示を生成条件として与えるものであり、選好をデータから推定して生成に接続する枠組みは想定されていない。

3 提案フレームワーク

本研究では、ペアワイズ比較データから推定した選好規則を用いてアイデア生成を制御する、指示文ベースのフレームワークを提案する。本フレームワークは追加学習やパラメータ更新を必要とせず、選好規則の推定と生成制御を同一の自然言語表現で接続する点に特徴がある。

3.1 フレームワーク概要

図1に示すように、提案フレームワークは、選好規則の推定と、推定規則に基づく生成からなる2段階構成である。ユーザは2つのアイデア候補のうち好ましい方を選択するのみであり、理由や基準の言語化は不要である。指示文例は §A に示す。

3.2 ペアワイズ比較による選好規則推定

入力は、2つのアイデア候補に対するユーザのペアワイズ比較結果である¹⁾。複数の比較事例を指示文中の例として与え、LLM に共通傾向を抽出させ、短い自然言語の集合として出力させる。推定される

1) ペアワイズ比較に用いるアイデア候補の生成方法や比較ペアの構成、ユーザフィードバック収集方法については、本稿では特定の手続きを仮定しない。

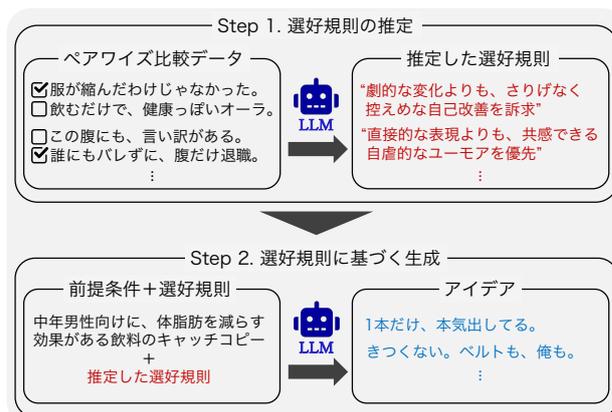


図1 提案フレームワークの概要。ペアワイズ比較データから LLM が選好規則を推定 (Step 1) し、前提条件と規則を与えて選好に沿うアイデアを生成する (Step 2)。

規則は、好ましい表現や特徴を簡潔に記述し、個々の比較を超えた選好傾向を表すことを目的とする。

推定後は任意の後処理として規則をスクリーニングし、データに基づく選別 [8] または人手による修正により選好規則を確定できる。

3.3 選好規則に基づくアイデア生成

生成段階では、生成タスクの前提条件 (対象や目的など) とともに選好規則を指示文に与え、LLM による生成を制御する。選好規則を生成条件として用いることで、一般的で無難な生成にとどまらず、個人や文脈に応じたアイデア生成を目指す。

4 実験

本章では、提案の有効性を検証するために行った2つの実験について述べる。実験には OpenAI の大規模言語モデル gpt-4.1-2025-04-14 を用いた。LLM のランダムネスによる出力の揺らぎを考慮し、評価結果は複数回実行した平均値として報告する。

4.1 実験 1：選好規則の推定と分類精度

本実験では、ペアワイズ比較データから推定した選好規則が、未知の比較に対する選好判別に有効かを評価する。具体的には、推定した選好規則を用いて、どちらのアイデアが好ましいかを判定させた分類精度を測定し、選好規則なしの設定と比較する。

4.1.1 データセット

人工的に定義した選好条件に基づき、2つの短文集合を LLM を用いて生成し、ペアワイズ比較データを構築した。表1に、選好条件の概要を示す。各

表1 実験1で用いた人工的な選好条件の一覧

条件名	説明 (左側を好ましいと定義)
topic	スポーツ／政治に関する短文
sentiment	ポジティブ／ネガティブな短文
brand	ブランド A / B 向けのキャッチコピー
target	若年層／高齢層向けのキャッチコピー
appeal	高級感／低価格訴求キャッチコピー
expression	意外性／分かりやすさ重視キャッチコピー

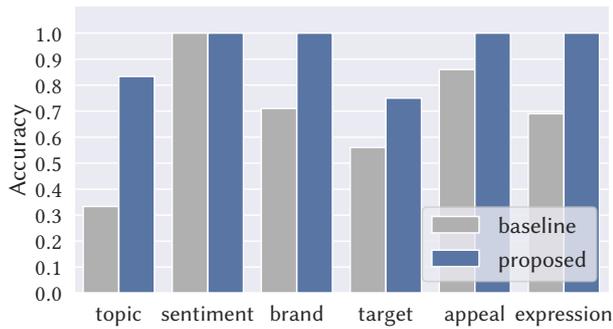


図2 実験1の結果：各選好条件における分類精度 (Accuracy)。

条件について20組の比較ペアを作成し、10組を選好規則推定用、残り10組を評価用として用いた。

4.1.2 評価方法

各条件について、推定用ペアを入力としてLLMに選好規則を推定させた。推定後、後処理として規則のスクリーニングを行い、採用する規則を確定した。具体的には、推定された各規則を単体で推定用ペアに適用し、規則が適用できた数と、正解ラベルを正しく予測できた数とを算出し、共に所定の閾値を上回る規則のみを採用した [8]。

評価用ペアに対し (A) 選好規則なしのベースライン (baseline)、(B) 推定した選好規則を与える提案手法 (proposed)、の2条件で選好判別を行った。

評価指標として、評価用ペアに付与された正解の選好ラベルと、モデルの判別結果が一致した割合を分類精度 (Accuracy) として用いた。

4.1.3 実験結果

図2に、各選好条件における分類精度を示す。提案手法では、多くの選好条件においてベースラインを上回る分類精度を示した。一方で、sentiment条件のように、ベースラインにおいても分類精度が1.0に達する条件が存在し、事前学習のみで選好を正しく判別可能な場合も見られた。この点については、§5.1で詳しく議論する。

表2 実験2で用いた課題の一覧

課題名	内容
bike	放置自転車問題に対する解決アイデア
cheat	定期テストにおけるカンニング防止策の提案
meeting	ミーティングの集まりを改善するための施策提案
night	夜間の帰宅路における安全確保策の提案
visitor	訪日外国人のための案内改善策の提案

4.2 実験2：選好規則に基づく生成評価

本実験では、ペアワイズ比較データから推定した選好規則を生成条件として用いた場合に、生成されたアイデアがより選好に適合するかを評価する。具体的には、同一の前提条件の下で、推定した選好規則を与えて生成したアイデアと、選好規則を与えずに生成したアイデアを比較し、どちらが好ましいかを判定する。本実験におけるペアワイズ比較データ、推定された選好規則、および生成アイデアの具体例は §Bに示す。

4.2.1 データセット

既存研究において、複数の課題に対するアイデアを、複数の観点ごとにクラウドワーカーが評価したデータセット [17] を用いた。本実験では、観点を選好条件として扱い、それに基づいてペアワイズ比較データを構築した。各課題について、評価傾向の類似性に基づいて抽出した代表的な観点を候補とし、比較ペアを安定に構成できる観点を選好条件として選択した。この方法の下で、5種類の課題を対象とし、各課題について5種類の選好条件を選択した。各 (課題, 選好条件) の組に対して、アノテーター間で平均した評価スコアに基づき10組の比較ペアを作成し、選好規則推定用とした。表2に、本研究で対象とした課題の概要を示す。

4.2.2 評価方法

各 (課題, 選好条件) について、実験1と同様に推定用ペアから選好規則を推定し、スクリーニングを経て採用規則を確定した。次に、(A) 選好規則を与えず前提条件のみを与えるベースライン、(B) 推定した選好規則を生成条件として与える提案手法、の2条件でそれぞれ10個のアイデアを生成した。

評価では、(A) と (B) の生成結果を1組として、真の選好条件を与えたLLM判定器に5段階で適合度を採点させた。得点を上から「適合」「中立」「不適合」の3段階に離散化し、異なる段階であれ

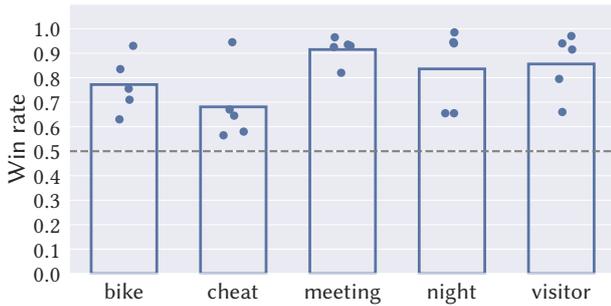


図3 実験2の結果：各課題における勝率（Win rate）

ば上位段階を勝者，同一段階であれば Tie として Win/Lose/Tie を判定した。勝率（Win rate）は，提案手法（B）を基準として，Win を 1.0，Lose を 0.0，Tie を 0.5 として平均を算出した。

なお，LLM 判定器の妥当性を確認するため，層化抽出²⁾した比較ペアに対し人手評価も行い，人手と LLM の Win/Lose/Tie 判定の一致率は 78%であった。

4.2.3 実験結果

図3に，各課題における勝率を示す。棒グラフは課題ごとの平均，点は各選好条件の結果，破線は（A）と（B）が同等な場合の期待値 0.5 を表す。いずれの（課題，選好条件）においても，提案手法の勝率は 0.5 を上回っており，ベースラインと比較して，提案手法によって生成されたアイデアはより選好に適合した。一方で，（課題，選好条件）によって勝率には差が見られ，条件によっては 0.6 程度と相対的に低い値を示す場合も確認された。

5 議論

5.1 特定の選好条件における性能低下

実験1では，sentiment 条件のように，選好規則を与えない設定においても分類精度が 1.0 に達するケースが確認された。さらに追加の検証として，同一の sentiment 条件において選好条件を反転させ，ネガティブな表現が好ましいと定義した場合には，選好規則を与えた場合・与えない場合のいずれにおいても分類精度が 0.0 となることが確認された。このとき，推定された選好規則は，意図した条件とは逆に，ポジティブな表現を好ましいとする内容を含んでいることも観察された。

このような挙動は，LLM が事前学習の過程で獲得した一般的なラベルバイアスの影響を強く受けて

2) 層は各（課題，選好条件）とし，各層から無作為に同数の比較ペア（各層 4 組，計 100 組）を抽出した。

いる可能性を示唆している。特に，文脈内学習において，与えられた少数の事例に含まれるラベル対応よりも，モデルが事前学習の過程で獲得したラベル間関係やラベル分布に基づく判断を優先してしまうバイアスにより，文脈内学習が十分に機能しない場合があることが指摘されている [18, 19, 20]。これらは文脈内学習における重要な課題であり，本研究においても今後の検討課題である。

5.2 データ量および提示順が与える影響

本研究では，各選好条件につき 10 組の推定用ペアを用い，提示順を固定して選好規則を推定した。文脈内学習では，提示する例の数やその順序がモデルの性能や判断に影響することが指摘されているが [15, 21, 22]，本研究ではこれらが規則の安定性に与える影響を十分に検証できていない。今後は，ペアワイズ比較データの組数や提示順を制御した実験により，提案の頑健性を検証する必要がある。

5.3 複数の選好が混在する状況での検証

本研究では，単一の選好条件に基づくペアワイズ比較データを用いた。一方，実際のアイデア生成支援の場面では，複数の選好基準が同時に存在し，状況に応じて使い分けられたり，互いに競合したりするケースが多い。このような複合的な選好構造に対して，提案フレームワークがどのような規則を推定し，複数の選好をどの程度分離または統合して表現できるかは，さらなる検証が必要である。

5.4 ユーザ実験による有効性の検証

本研究では，人工的に設計した選好条件や既存データセットに基づくペアワイズ比較データを用いて，提案の有効性を検証した。今後は，ユーザ実験を通じて，実際の利用場面における選好の多様性や曖昧さを踏まえた上で，提案がアイデア生成支援にどの程度有用であるかを検証する必要がある。

6 おわりに

本研究では，ペアワイズ比較データから LLM が自然言語で選好規則を推定し，その規則を生成条件として直接利用する統合フレームワークを提案した。実験により，推定した選好規則が選好判別および生成タスクの双方において有効であることを示した。今後は，実ユーザを対象とした評価を通じて，実用的有効性の検証を進める。

謝辞

本論文の執筆にあたり実験をサポートいただいたデータ・アナリティクス・ラボ株式会社の阿左美拓音様に感謝いたします。

参考文献

- [1] Srishti Palani and Gonzalo Ramos. Evolving roles and workflows of creative practitioners in the age of generative AI. In **Proc. C&C**, 2024.
- [2] Jessica He, Stephanie Houde, Gabriel E Gonzalez, Darío Andrés Silva Moran, Steven I Ross, Michael Muller, and Justin D Weisz. AI and the future of collaborative work: Group ideation with an LLM in a virtual canvas. In **Proc. CHIWORK**, 2024.
- [3] Orit Shaer, Angelora Cooper, Osnat Mokryn, Andrew L Kun, and Hagit Ben Shoshan. AI-augmented brainwriting: Investigating the use of LLMs in group ideation. In **Proc. CHI**, 2024.
- [4] Anil R Doshi and Oliver P Hauser. Generative artificial intelligence enhances creativity but reduces the diversity of novel content. **Sci. Adv.**, Vol. 10, No. 28, 2024.
- [5] Noor Hammad, C Ailie Fraser, Erik Harpstead, Jessica Hammer, and Mira Dontcheva. “It’s more of a vibe I’m going for”: Designing text-to-music generation interfaces for video creators. In **Proc. DIS**, 2025.
- [6] Yiyang Xu, Jinghao Zhang, Alireza Salemi, Xinting Hu, Wenjie Wang, Fuli Feng, Hamed Zamani, Xiangnan He, and Tat-Seng Chua. Personalized generation in large model era: A survey. In **Proc. ACL**, 2025.
- [7] Dongyoung Go, Tomasz Korbak, Germán Kruszewski, Jos Rozen, and Marc Dymetman. Compositional preference models for aligning LMs. In **Proc. ICLR**, 2024.
- [8] Arduin Findeis, Timo Kaufmann, Eyke Hüllermeier, Samuel Albanie, and Robert Mullins. Inverse constitutional AI: Compressing preferences into principles. In **Proc. ICLR**, 2025.
- [9] Paul Christiano, Jan Leike, Tom B Brown, Miljan Martić, Shane Legg, and Dario Amodei. Deep reinforcement learning from human preferences. In **Proc. NeurIPS**, 2017.
- [10] Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. Training language models to follow instructions with human feedback. In **Proc. NeurIPS**, 2022.
- [11] Rafael Rafailov, Archit Sharma, Eric Mitchell, Stefano Ermon, Christopher D Manning, and Chelsea Finn. Direct preference optimization: Your language model is secretly a reward model. In **Proc. NeurIPS**, 2023.
- [12] Sriyash Poddar, Yanming Wan, Hamish Ivison, Abhishek Gupta, and Natasha Jaques. Personalizing reinforcement learning from human feedback with variational preference learning. In **Proc. NeurIPS**, 2024.
- [13] Nitish Shirish Keskar, Bryan McCann, Lav R Varshney, Caiming Xiong, and Richard Socher. CTRL: A conditional transformer language model for controllable generation. **arXiv [cs.CL]**, 2019.
- [14] Sumanth Dathathri, Andrea Madotto, Janice Lan, Jane Hung, Eric Frank, Piero Molino, Jason Yosinski, and Rosanne Liu. Plug and play language models: A simple approach to controlled text generation. In **Proc. ICLR**, 2020.
- [15] Tom B Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners. In **Proc. NeurIPS**, 2020.
- [16] Jason Wei, Maarten Bosma, Vincent Y Zhao, Kelvin Guu, Adams Wei Yu, Brian Lester, Nan Du, Andrew M Dai, and Quoc V Le. Finetuned language models are zero-shot learners. In **Proc. ICLR**, 2022.
- [17] Yukino Baba, Jiyi Li, and Hisashi Kashima. CrowDEA: Multi-view idea prioritization with crowds. In **Proc. HCOMP**, 2020.
- [18] Sewon Min, Xinxin Lyu, Ari Holtzman, Mikel Artetxe, Mike Lewis, Hannaneh Hajishirzi, and Luke Zettlemoyer. Rethinking the role of demonstrations: What makes in-context learning work? In **Proc. EMNLP**, 2022.
- [19] Jannik Kossen, Yarin Gal, and Tom Rainforth. In-context learning learns label relationships but is not conventional learning. In **Proc. ICLR**, 2024.
- [20] Yu Fei, Yifan Hou, Zeming Chen, and Antoine Bosselut. Mitigating label biases for in-context learning. In **Proc. ACL**, 2023.
- [21] Tony Z Zhao, Eric Wallace, Shi Feng, Dan Klein, and Sameer Singh. Calibrate before use: Improving few-shot performance of language models. In **Proc. ICML**, 2021.
- [22] Yao Lu, Max Bartolo, Alastair Moore, Sebastian Riedel, and Pontus Stenetorp. Fantastically ordered prompts and where to find them: Overcoming few-shot prompt order sensitivity. In **Proc. ACL**, 2022.

A 指示文

Listing 1に選好規則推定の際に用いる指示文例を示す。また、Listing 2にアイデア生成の際に用いる指示文例を示す。

Listing 1 選好規則推定のための指示文例

```
You are given a list of paired data points.
Each pair (x_0, x_1) represents two items.

For each pair, we know which item was selected based on
an unknown rule. Your task is to infer
the classification rule used to generate these labels.

To do this, carefully compare each item in the pair and
identify specific features — such as content, topic,
tone, clarity, structure, or wording — that
distinguish the selected item from the other.
Focus on concise, consistent patterns across pairs
that could explain the selections.

Each line is formatted as:
pair_i: x_0 = <value_0>, x_1 = <value_1>,
      label = <0 or 1>
where:
- label = 0 means x_0 is selected over x_1
- label = 1 means x_1 is selected over x_0

[Training Data - For Rule Inference]
{training_data}

Your output should be in the following JSON format:
- analysis_process: Describe the process by which
  you conducted your analysis here.
- hypothesis: Conclude with a list of short, specific
  rules — each 10 words or fewer —
  that explain the selections observed in the data.

Important:
- List only unique rules; do not repeat
  or rephrase similar ones.
- Merge semantically identical ideas into
  a single rule.
- Keep each rule short, specific, and self-contained.
- Do not include any additional explanation
  or formatting outside of the JSON object.

Language: Japanese
```

Listing 2 選好規則に基づくアイデア生成の指示文例

```
The following is a problem to be solved:
{problem}

Please generate {n} concise ideas that address
this problem, focusing on the following viewpoint:
{inferred_rule}

Each idea should be 1 to 2 plain sentences.
List them one per line, without numbering
or explanations.
```

B 実験データ具体例

本節では、実験2の放置自転車問題 (bike) において「罰則として違反者にダイレクトに響く」という観点 (選好条件) を対象に、データ例を一部示す。

B.1 ペアワイズ比較データ

- A: 自転車を放置した場合、罰金を徴収する
B: 地面に絵を描き、地面にデザイン性をもたせる
(選択: A)
- A: 警備員を巡回させ、放置する人を見つけたら罰金をとるようにする
B: 棟ごとに駐輪スペースを設ける
(選択: A)
- A: 自転車が放置される場所の地面に立体アートを描く
B: 放置自転車を全て取り除き罰則を与える
(選択: B)

B.2 推定された選好規則

- 罰則や強制措置を含む案が優先される
- 具体的な行動や仕組みが明記された案が選ばれる
- 啓蒙や美化のみの案は選ばれにくい

B.3 生成されたアイデア

- 規則なし (baseline)
 - 放置自転車専用の回収日を設けて定期的に撤去する。
 - 大学敷地内に監視カメラを設置し、放置を抑止する。
 - 自転車利用者に登録制ステッカーを配布して所有者管理を徹底する。
- 規則あり (proposed)
 - 定期的にキャンパス内で放置自転車の撤去作業を実施し、違反者には罰金を科す。
 - 自転車駐輪許可証制度を導入し、無許可自転車は発見次第撤去および所有者に罰則を科す。
 - 放置自転車には警告札を取り付け、一定期間以内に移動しない場合は強制回収する。