

# 招待論文 CatAlyst: 大規模生成モデルを用いた知的タスクの先延ばし防止インタラクション

荒川 陸<sup>\*†</sup> 矢倉 大夢<sup>\*‡</sup> 後藤 真孝<sup>§</sup>

**概要.** 本論文では、大規模生成モデルを用いて知的タスクの「Procrastination (先延ばし)」を減らすインタラクションシステム CatAlyst [4] について述べる。CatAlyst は先延ばししようとしているタスクへのエンゲージメントを高めるべく、中断した作業内容の続きを生成モデルで生成して提示する。これは、タスクを中断したユーザに対して事前に定められたメッセージを提示する既存手法に比べ、ユーザのコンテキストを踏まえた介入となり、タスクの再開を促す効果を持つ。また、モデルの生成物が十分な質でなかった場合でも、タスクを再開することへのハードルを下げることで、ユーザのタスク遂行に貢献すると期待できる。こうした点について、文書執筆とスライド編集の2つのタスクで CatAlyst を用いたユーザ実験を行い、認知負荷を軽減させながらタスクの再開を促進する効果を確認した。これは、大規模生成モデルを個々のタスクに合わせてチューニングせずとも、ユーザの行動変容を促す介入として用いることで、多様なタスクでユーザの知的生産活動を支援できる可能性を示唆するものである。

## 1 はじめに

近年の大規模生成モデルの進展により、文書執筆やスライド作成などの知的タスクの一部を AI に代替する試みが増えている [25, 35]。しかし、個々のタスクにモデルをチューニングするコストが膨大であること [26, 1] や、生成結果が必ずしもユーザの期待する質に達しないこともあるという点 [11, 15, 32, 3] から十分な利用が進んでいない部分もある。

一方で、こうした知的タスクは長時間の集中を要することも多く [22, 23, 27]、ユーザが途中で意欲をなくして「Procrastination (先延ばし)」してしまうという課題もある [18, 13, 5]。このようなユーザのタスクへのエンゲージメントを促進する既存のアプローチとして、ユーザを励ますメッセージを提示するなどの介入が提案されているものの [6, 10, 20]、必ずしもユーザの作業のコンテキストを反映したのではなく、また提示内容のバリエーションにも欠くため、改善の余地があるものと考えられる [24, 17]。

そこで本研究 [4] では、大規模生成モデルをチューニングなしに利用しながら、ユーザの行動変容を促し知的タスクの先延ばしを防止する新たなインタラクション「CatAlyst」を提案した。CatAlyst は、ユーザがタスクを先延ばししようとしている時に、中断した作業内容の続きを生成モデルを用いて提示することで、再開のハードルを下げることを目指す。

ユーザ実験では、文書執筆とスライド編集の2つのタスクで CatAlyst の有効性を評価しており、それらの結果からチューニングなしに既存の大規模生成モデルを幅広いタスクにおけるユーザの知的生産活動支援に応用できる可能性が示唆された。

## 2 着想

本研究の端緒は、先延ばしの対策を探していた著者らの経験に遡ることができる。試行錯誤の末の結論としては、やる気を出す一番の方法は「とにかく作業を始めること」にあると思われた。そして、どのようにすればインタラクションの力でその最初の一步を踏み出しやすくできるかを検討したことが、本研究につながった。具体的には、他人の作業内容に対するレビューであれば<sup>1</sup>、仮にそれが完璧でなかったとしても、自分で0から作り始めるのに比べて作業開始の負荷が小さくなる傾向があるという観察に着想を得た。これは、後述する Fogg のモデル [12] とも関連している。

## 3 提案アプローチ

CatAlyst はまず、ユーザのタスクの進捗が停滞したタイミングを、操作ログ等から検出する。次に、中断した時点までの作業内容を入力として大規模生成モデルに与え、その続きを生成させる。例えば、文書執筆においてはユーザがそれまでに書いた文章を入力として、またスライド作成においてはそれまでに作成したスライドの内容をプロンプトへと加工

Copyright is held by the author(s). This paper is nonrefereed and non-archival.

\* Authors contributed equally.

† Carnegie Mellon University

‡ 筑波大学

§ 産業技術総合研究所

<sup>1</sup> これは、筆頭著者の2名が共同で複数の論文を執筆してきた所以の1つである。

表 1: ユーザ実験において、コントロール条件に比べ CatAlyst の導入が有意な変化を与えた項目

評価項目	概略	文書作成	スライド作成
<b>Ignorance rate</b>	通知が無視された割合	✓	✓
<b>Interest retrieval time</b>	タスク再開までに要した時間	✓	✓
<b>Progress after resumption</b>	タスク再開後一定時間内の作業量	✓	✓
<b>Total time</b>	タスク全体の所要時間	有意差なし	✓
<b>Subjective quality</b>	第三者による成果物の評価	有意差なし	✓
<b>Cognitive load</b>	参加者による NASA-TLX [14, 8] の評価	✓	✓
<b>System usability</b>	参加者による SUS [7] の評価	✓	✓

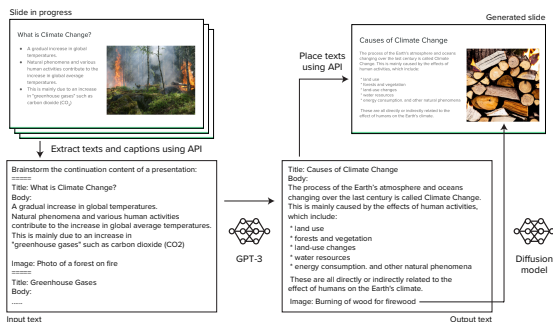


図 1: スライド作成タスクにおける生成の流れ

して (図 1), それらに続く内容を生成させる. そして, 生成された内容の一部をユーザに通知として提示することで, タスクの再開を促す.

これは, ユーザの行動変容は「Motivation (動機)」「Ability (実行能力)」「Prompt (きっかけ)」の組み合わせによって引き起こされるとした Fogg のモデル [12] に依拠したものである. 具体的には, 事前に定められたメッセージを提示するのではなく, 中断したタスクの状況に合わせた内容を動的に提示することで, タスクの再開への動機を引き出す. さらに, CatAlyst が AI システムとして共同でタスクに取り組んでいるという認識を作ることができれば, さらなる動機づけにつながると考えられる [34, 33]. また, 作業内容の続きを生成して提示することで, ユーザがその一部を成果物に取り入れたり, 方向性のアイデアを得たりといった形で, タスクに取り組むこと自体がより容易になる可能性もある. ここで重要なのは, 必ずしも生成結果の質が完璧である必要はなく, そのため既存の大規模生成モデルを様々な知的タスクへと援用することが可能だという点である. そして, 生成内容をユーザに通知し, 作業を再開させるきっかけを与えることで, タスクの再開にかかる「活性化エネルギー [9]」を低減させる.

## 4 ユーザ実験

本研究では, CatAlyst の有効性を文書執筆とスライド作成という 2 種類の知的タスクにおいて検証した. それぞれ 12 名の参加者が, 介入のないベースライン条件と, タスク中断時に予め定められたメッ

セージから 1 つをランダムに提示するコントロール条件と比較する形で, 参加者内実験として CatAlyst を利用した. 結果は表 1 に示す通り, 幅広い評価項目について CatAlyst の導入が有意な変化を与えており, ユーザの知的生産活動への貢献という点でその効果が示唆された. また, 中長期的な導入がユーザの知的生産活動に与える影響も調査すべく, CatAlyst を文書執筆に自由に利用できる環境を, 10 名の参加者に 5 日間提供し, その使用状況を調査するというも行った. 提供期間後の半構造化インタビューからは, 参加者がそれぞれ異なる形で CatAlyst を利用している様子も明らかになった.

## 5 議論

近年の Human-AI Collaboration に関する研究の多くは, 人間が取り組んできたタスクのより多くを自動化することを目指している [21] が, その結果が人間の期待に満たないときに信頼を損なってしまうという問題も報告されてきた [30, 16, 29]. それらに対し本研究では, 生成モデルをユーザの作業ルーチンの一部として取り込むのではなく, 作業から離れたときの補助的な介入のために限定的に用いるというアプローチをあえて採っている. これにより, 生成モデルへの期待が「作業を代替してくれるもの」という位置づけではなくなり, その出力の質が不十分であってもユーザの信頼や体験を損ないにくいという示唆が得られた. このアプローチは, Yang らのいう「Unremarkable AI (目立たない AI)」 [31] と関連付けることができ, AI システムがあえてその存在感を制限するというインタラクションデザインの可能性を示している. 一方, 長期的な利用を行った参加者からは, 能動的に生成モデルを利用して作業内容の続きを生成するといったユースケースについての要望も得られた. その点で例えば, オートコンプリート [19] や GitHub Copilot [35] のような形で, 受動性と能動性のバランスを取ったインタラクションを探索することは有効であると考えられる.

## 謝辞

本研究は JST ACT-X (JPMJAX200R) 及び JSPS 科研費 (JP21J20353) の助成を受けている.

## 参考文献

- [1] L. F. W. Anthony, B. Kanding, and R. Selvan. Carbontracker: Tracking and Predicting the Carbon Footprint of Training Deep Learning Models. In *Proceedings of the 2020 ICML Workshop on Challenges in Deploying and monitoring Machine Learning Systems*, Ithaca, NY, 2020. arXiv.
- [2] R. Arakawa and H. Yakura. INWARD: A Computer-Supported Tool for Video-Reflection Improves Efficiency and Effectiveness in Executive Coaching. In *Proceeding of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, New York, NY, 2020. ACM.
- [3] R. Arakawa and H. Yakura. AI for human assessment: What do professional assessors need? *arXiv*, abs/2204.08471, 2022.
- [4] R. Arakawa, H. Yakura, and M. Goto. CatAlyst: Domain-Extensible Intervention for Preventing Task Procrastination Using Large Generative Models. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 157:1–157:19, New York, NY, 2023. ACM.
- [5] M. Beheshtifar, H. Hoseinifar, and M. N. Moghadam. Effect Procrastination on Work-Related Stress. *European Journal of Economics, Finance and Administrative Sciences*, 38(38):59–64, 2011.
- [6] A. M. Z. Bernuy, Z. Han, H. Shaikh, Q. Y. Zheng, L. Lim, A. N. Rafferty, A. Petersen, and J. J. Williams. How can Email Interventions Increase Students’ Completion of Online Homework? A Case Study Using A/B Comparisons. In *Proceedings of the 12th International Learning Analytics and Knowledge Conference*, pp. 107–118, New York, NY, 2022. ACM.
- [7] J. Brooke. SUS: A ‘Quick and Dirty’ Usability Scale. In P. W. Jordan, B. Thomas, I. L. McClelland, and B. Weerdmeester eds., *Usability Evaluation In Industry*, pp. 207–212. CRC Press, London, UK, 1996.
- [8] J. C. Byers, A. C. Bittner, and S. G. Hill. Traditional and Raw Task Load Index (TLX) Correlations: Are Paired Comparisons Necessary? In *Proceedings of the 1989 Annual International Industrial Ergonomics and Safety Conference*, pp. 481–485, Philadelphia, PA, 1989. Taylor & Francis.
- [9] J. Clear. The Chemistry of Building Better Habits, 2015. <https://jamesclear.com/chemistry-habits>. Last accessed on September 2022.
- [10] R. A. J. de Vries, C. Zaga, F. Bayer, C. H. C. Drossaert, K. P. Truong, and V. Evers. Experts Get Me Started, Peers Keep Me Going: Comparing Crowd- Versus Expert-Designed Motivational Text Messages for Exercise Behavior Change. In *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare*, pp. 155–162, New York, NY, 2017. ACM.
- [11] B. J. Dietvorst, J. P. Simmons, and C. Massey. Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err. *Journal of Experimental Psychology: General*, 144(1):114–126, 2015.
- [12] B. J. Fogg. A Behavior Model for Persuasive Design. In *Proceedings of the 4th International Conference on Persuasive Technology*, No. 40, pp. 1–7, New York, NY, 2009. ACM.
- [13] M. D. Frakes and M. F. Wasserman. Procrastination in the Workplace: Evidence from the US Patent Office. Technical report, National Bureau of Economic Research, 2016.
- [14] S. G. Hart and L. E. Staveland. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology*, 52:139–183, 1988.
- [15] E. Jussupow, I. Benbasat, and A. Heinzl. Why Are We Averse Towards Algorithms? A Comprehensive Literature Review on Algorithm Aversion. In *Proceedings of the 28th European Conference on Information Systems*, pp. 1–16, Atlanta, GA, 2020. AIS.
- [16] J. Kim, M. L. Maher, and S. Siddiqui. Studying the Impact of AI-based Inspiration on Human Ideation in a Co-Creative Design System. In *Proceedings of the 2nd ACM IUI Workshops on Human-AI Co-Creation with Generative Models*, Aachen, Germany, 2021. CEUR-WS.org.
- [17] G. Kovacs, Z. Wu, and M. S. Bernstein. Rotating Online Behavior Change Interventions Increases Effectiveness But Also Increases Attrition. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW):95:1–95:25, 2018.
- [18] C. H. Lay. At Last, My Research Article on Procrastination. *Journal of Research in Personality*, 20(4):474–495, 1986.
- [19] F. Lehmann and D. Buschek. Examining Autocompletion as a Basic Concept for Interaction with Generative AI. *i-com*, 19(3):251–264, 2021.
- [20] Y. Liu, Y. Jia, W. Pan, and M. S. Pfaff. Supporting Task Resumption Using Visual Feedback. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work*, pp. 767–777, New York, NY, 2014. ACM.
- [21] B. Lubars and C. Tan. Ask Not What AI Can Do, but What AI Should Do: Towards a Framework of Task Delegability. In *Proceedings of the 2019 Annual Conference on Neural Information Processing Systems*, pp. 57–67, Red Hook, NY, 2019. Curran Associates.
- [22] N. H. Mackworth. The Breakdown of Vigilance During Prolonged Visual Search. *Quarterly Journal of Experimental Psychology*, 1(1):6–21, 1948.
- [23] R. Parasuraman. Memory Load and Event Rate Control Sensitivity Decrements in Sustained Attention. *Science*, 205(4409):924–927, 1979.

- [24] J. A. Rodriguez, G. Piccoli, and M. Bartosiak. Nudging the Classroom: Designing a Socio-Technical Artifact to Reduce Academic Procrastination. In *Proceedings of the 52nd Hawaii International Conference on System Sciences*, pp. 1–10, Honolulu, HI, 2019. ScholarSpace.
- [25] T. Schick, J. Dwivedi-Yu, Z. Jiang, F. Petroni, P. Lewis, G. Izacard, Q. You, C. Nalmpantis, E. Grave, and S. Riedel. PEER: A Collaborative Language Model. *arXiv*, 2208.11663, 2022.
- [26] E. Strubell, A. Ganesh, and A. McCallum. Energy and Policy Considerations for Modern Deep Learning Research. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence*, pp. 13693–13696, Palo Alto, CA, 2020. AAAI Press.
- [27] B. Wang, Y. Liu, J. Qian, and S. K. Parker. Achieving Effective Remote Working During the COVID-19 Pandemic: A Work Design Perspective. *Applied Psychology*, 70(1):16–59, 2020.
- [28] H. Yakura, Y. Koyama, and M. Goto. Tool- and Domain-Agnostic Parameterization of Style Transfer Effects Leveraging Pretrained Perceptual Metrics. In *Proceedings of the 30th International Joint Conference on Artificial Intelligence*, pp. 1208–1216, Palo Alto, CA, 2021. AAAI Press.
- [29] C. Yan, J. J. Y. Chung, Y. Kiheon, Y. I. Ginzgold, E. Adar, and S. R. Hong. FlatMagic: Improving Flat Colorization through AI-driven Design for Digital Comic Professionals. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pp. 380:1–380:17, New York, NY, 2022. ACM.
- [30] D. Yang, Y. Zhou, Z. Zhang, T. J. Li, and R. LC. AI as an Active Writer: Interaction Strategies with Generated Text in Human-AI Collaborative Fiction Writing. In *Proceedings of the 3rd ACM IUI Workshops on Human-AI Co-Creation with Generative Models*, pp. 56–65, Aachen, Germany, 2022. CEUR-WS.org.
- [31] Q. Yang, A. Steinfeld, and J. Zimmerman. Unremarkable AI: Fitting Intelligent Decision Support into Critical, Clinical Decision-Making Processes. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, p. 238, New York, NY, 2019. ACM.
- [32] M. Yin, J. W. Vaughan, and H. M. Wallach. Understanding the Effect of Accuracy on Trust in Machine Learning Models. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, p. 279, New York, NY, 2019. ACM.
- [33] C. Zhang, C. Yao, J. Wu, W. Lin, L. Liu, G. Yan, and F. Ying. StoryDrawer: A Child-AI Collaborative Drawing System to Support Children’s Creative Visual Storytelling. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pp. 311:1–311:15, New York, NY, 2022. ACM.
- [34] C. Zhang, Z. Zhou, J. Wu, Y. Hu, Y. Shao, J. Liu, Y. Hu, F. Ying, and C. Yao. Bio Sketchbook: An AI-assisted Sketching Partner for Children’s Biodiversity Observational Learning. In *Proceedings of the 20th ACM Interaction Conference on Design and Children, Athens*, pp. 466–470, New York, NY, 2021. ACM.
- [35] A. Ziegler, E. Kalliamvakou, X. A. Li, A. Rice, D. Rifkin, S. Simister, G. Sittampalam, and E. Aftandilian. Productivity Assessment of Neural Code Completion. In *Proceedings of the 6th ACM SIGPLAN International Symposium on Machine Programming*, pp. 21–29, New York, NY, 2022. ACM.

## 未来ビジョン

ここ数年、大規模言語モデルを含めた大規模マルチモーダルモデルの圧倒的な能力がある中で、多くの企業や研究チームがその学習に資金を投じてきた。しかし、これらモデルの学習は莫大なコストのみならず環境への負荷ももたらしており、各社が試行錯誤を繰り返している先に、学習のコストや負荷が増大していくことが予見される。こうした状況において、サステイナブルにモデルを活用していくビジョンを描き、その普遍的な方法論を残していくということが、これから先の未来において研究者に求められていくことだろう。

その中で我々は、様々な応用シナリオごとに精度を上げるためのコスト（例えば fine-tuning）をさらに費やすのではなく、既存のモデルそのままでもユーザの役に立つようにインタラクションデザインを工夫するという、

HCI 研究者ならではのアプローチを本研究以外を含め提示してきた [28]。また、そうしたアプローチの延長上に、モデルの不完全性を許容するインタラクションデザインについての提案も行っている。例えば、ACM CHI '21において我々は、過去にも不完全なセンシングモデルに基づく介入手法のデザイン研究を発表しており [2]、false positive がユーザの信頼や体験を損なわないようなデザインを実現することが、サステイナブルな Human-AI Collaboration を実現する上で重要であることを議論した。大規模モデルによって次々に新しいことが可能になる時代だからこそ、その中でも変わらない問題を考え続けていきたい。