

### Adding Conditional Control to Diffusion Models with Reinforcement Learning

**Yulai Zhao**\*, Masatoshi Uehara\*, Gabriele Scalia, Sunyuan Kung, Tommaso Biancalani, Sergey Levine, Ehsan Hajiramezanali

#### Acknowledgements

- Accepted at ICLR 2025. Paper is released at <u>https://arxiv.org/abs/2406.12120</u>
- GitHub repo: <u>https://github.com/zhaoyl18/CTRL</u>
- This work was done when Yulai Zhao was a research intern at Genentech.

#### General roadmap: AI-aided design



#### **Examples - Images**



In this work, our objective is different from merely optimizing towards certain reward functions!

#### **Standard fine-tuning: towards reward model**

![](_page_5_Picture_1.jpeg)

![](_page_6_Picture_0.jpeg)

#### **Examples – Sequence design**

![](_page_7_Figure_1.jpeg)

cell line

## Why our task is important?

Condition on high activity level in HepG2

![](_page_8_Figure_2.jpeg)

Also active in other cell lines, no specificity

Add "conditional control"

![](_page_8_Figure_5.jpeg)

Can achieve specificity!

![](_page_9_Figure_0.jpeg)

![](_page_10_Figure_0.jpeg)

![](_page_10_Picture_1.jpeg)

#### How to train diffusion models?

![](_page_11_Figure_1.jpeg)

#### Our goal: adding control via fine-tuning

![](_page_12_Figure_1.jpeg)

#### Our goal: adding control via fine-tuning

![](_page_13_Figure_1.jpeg)

#### Methodology

![](_page_14_Figure_1.jpeg)

#### **Reward collapse**

![](_page_15_Picture_1.jpeg)

- 1. Very common in generative systems (DMs, LLMs, GANs)
- 2. Generations have low diversity
- 3. Because the oracle is "over-optimized"

#### Methodology – Cont.

![](_page_16_Figure_1.jpeg)

![](_page_17_Figure_0.jpeg)

#### **Advantage: conditional independency**

![](_page_18_Figure_1.jpeg)

#### Experiments

![](_page_19_Figure_1.jpeg)

#### Compare with baselines

- Classifier guidance: training a classifier and incorporating its gradients to guide inference (while freezing pre-trained models)
- Classifier-free guidance: directly conditions the generative process on both data and context, bypassing the need for explicit classifiers.

Table 1: Comparison between our proposal and existing approaches. In contrast to classifier guidance or its variations, our method entails re-training the models directly on top of pre-trained models (i.e., fine-tuning). Additionally, we circumvent the necessity of learning a mapping  $x_t \rightarrow y$  or employing heuristic approximation techniques to address this issue. Compared to classifier-free guidance which always demands triplets  $\{c, x, y\}$ , our method can leverage conditional independence and only necessitate pairs  $\{x, y\}$  by leveraging if  $y \perp c | x$  holds. This simplifies the construction of the offline dataset.

Methods	Fine-tuning	Need to learn $x_t \to y$	Leveraging conditional independence
Classifier guidance (Dhariwal and Nichol, 2021)	No	Yes	Yes
Reconstruction guidance (e.g. (Ho et al., 2022), (Chung et al., 2022), (Han et al., 2022))	No	No	Yes
Classifier-free guidance (Ho and Salimans, 2022)	Yes	No	No
CTRL (Ours)	Yes	No	Yes

## **Example 1: Compressibility**

![](_page_21_Figure_1.jpeg)

	Accuracy ↑	Macro F1 score ↑	
DPS	0.45	0.44	Y=3
CTRL (Ours)	1.0	1.0	

(b) Evaluation of conditional generations

![](_page_21_Figure_4.jpeg)

(c) Generated images

more compressible

#### **More Qualitative Results**

![](_page_22_Figure_1.jpeg)

Figure 4: More images generated by CTRL in the compressibility task.

#### **Example 2: Compressibility & Aestheticness**

![](_page_23_Figure_1.jpeg)

#### **More Qualitative Results**

![](_page_24_Figure_1.jpeg)

Figure 5: More images generated by CTRL in the multi-task conditional generation.

## Extension: ada ding condition on continuous y

![](_page_25_Figure_1.jpeg)

## Conclusion

- We introduce an RL-based fine-tuning approach for conditioning pretrained diffusion models on new additional labels.
- Compared to classifier-free guidance, our proposed method uses the offline dataset more efficiently and allows for leveraging the conditional independence assumption, thereby greatly simplifying the construction of the offline dataset.
- We also theoretically justify our approach and build the connection with classifier-based guidance.

## **Future work**

- Extending this work to biological sequence design, such as DNA enhancers and RNA 5'UTR.
- The high-level goal is to achieve cell-specific promoter design.

# Thank you!