Legibility Diffuser: Offline Imitation for Intent Expressive Motion

Matthew Bronars, Shuo Cheng, Danfei Xu

***Abstract*— In human-robot collaboration, legible motion that conveys a robot's intentions and goals is known to improve safety, task efficiency, and user experience. Legible robot motion is typically generated using hand-designed cost functions and classical motion planners. However, with the rise of deep learning and data-driven robot policies, we need methods for training end-to-end on offline demonstration data. In this paper, we propose Legibility Diffuser, a diffusion-based policy that learns intent expressive motion directly from human demonstrations. By variably combining the noise predictions from a goal-conditioned diffusion model, we guide the robot's motion toward the most legible trajectory in the training dataset. We find that decaying the guidance weight over the course of the trajectory is critical for maintaining a high success rate while maximizing legibility.**

I. INTRODUCTION

Imitation learning (IL) is a powerful paradigm that allows robot policies to be trained on previously collected human demonstrations. Offline IL allows robotics to scale with big data, eliminating the need for costly environment interaction. When training robots for human environments, leveraging offline IL is especially important for safety and effectiveness. For this reason, developing learning from demonstrations (LfD) algorithms that are amenable to human-robot interaction (HRI) is an important avenue of research. One important characteristic of cooperative robots is legible motion that clearly conveys the robot's intentions and goals. It then seems natural to ask: *How can we directly learn legible robot motion from previously collected human demonstrations?*

As robots become more integrated into our daily lives, it is critical that they move in a way that is not only efficient and functional, but also legible and understandable to humans. In HRI, legible motion conveys a robot's intentions and goals in an intuitive and interpretable manner [1], [2]. Making a robot's actions more transparent will allow humans to better anticipate and respond to the robot. This can reduce the risk of accidents and collisions, which is important in safety critical environments. Concretely, legible robot motion allows an observer to make early and accurate predictions of an agent's target goal. Studies have shown that in collaborative environments, this leads to faster task completion times and fluent collaboration [3], [4].

Mathematically, a legible trajectory is one that maximizes $p(g^*|\xi_{s_0 \to s_t})$ where g^* is the goal and $\xi_{s_0 \to s_t}$ is the ongoing trajectory [1]. Methods for generating legible motion traditionally leverage hand designed cost functions and classical motion planning algorithms to maximize this term [1]. These classical approaches lack the scalability and the flexibility afforded by deep imitation learning algorithms. Deep learning approaches are state of the art in natural language processing [5] and computer vision [6]. Large scale data collection efforts are paving the way for these data driven approaches in robotics [7]. In order to reap the benefits of these large scale datasets, we need methods for legible motion generation that learn directly from human demonstrations.

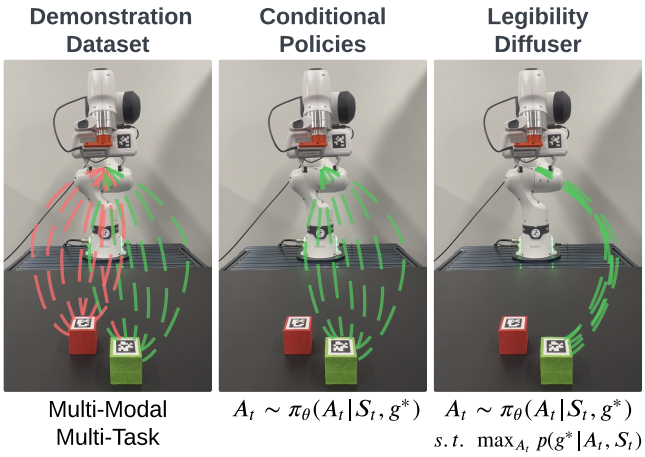

Fig. 1. Legibility Diffuser is an off-line imitation learning algorithm that can learn to generate the most legible modes from a multi-modal, multi-task dataset of human demonstrations.

To this end, our paper establishes a connection between legible motion and conditional generative models [8], [9], [10]. A critical feature of realistic human demonstrations is that they are multi-task and multi-modal [11], [12]; they show multiple ways of accomplishing a task or reaching a goal. Our key insight is that we can generate intent expressive motion by directly imitating the most legible mode from a diverse dataset of human demonstrations. Specifically, we introduce Legibility Diffuser (Fig 1), a conditional generative policy that produces legible robot motion through diffusion model guidance [13]. Our end-to-end method does not require estimating cost functions, classical motion planning, or any labeling of the training dataset.

Our results show that Legibility Diffuser is able to successfully imitate the most legible trajectories from a multi-modal and multi-task demonstration dataset. Compared to other imitation learning baselines, we find that Legibility Diffuser is statistically significantly more legible across four experiments, including a long-horizon manipulation task, and

*This work was not supported by any organization

[1]Matthew Bronars, Shuo Cheng, and Danfei Xu are with the school of interactive computing, Georgia Institute of Technology, 30332 Atlanta GA, United States of America. mbronars@gatech.edu, shuocheng@gatech.edu, danfei@gatech.edu

a real robot evaluation. We are able to accomplish this without access to the hand design cost functions that are typically required to generate legible motion. Our experiment shows that in the action generation domain, the class specific features that are enhanced by diffusion model guidance are the same features that make a trajectory legible. So just as increasing guidance weights improves the quality of image generators [13], increasing guidance weights improves the *legibility* of action generators. We also find that decaying the guidance weight over the course of the trajectory is important for maintaining a high success rate.

II. PREVIOUS WORK

A. Legible Robot Motion

Shared intentionality is an important aspect of human cognition, and being able to read intentions is critical for how we collaborate as a species [14]. Intent expressive actions (legible actions) are a form of non-verbal communication that allows groups of agents to coordinate their behaviors. This is useful for HRI because if robots forecast their next move, they can fluidly interact and improvise with humans [15]. Robots are more readable and understandable if they have the capability to express forethought and respond to task outcomes. This increases people's perception of robots and will make users more willing to engage in interactions with legible robots [16]. Experiments have shown that legible motion allows for faster completion time of collaborative tasks and increased user satisfaction [4]. Importantly, motion produced by robots is legible if it allows for quick and confident predictions of the goal state.

Standard methods for generating legible motion [1] involve hand designed cost functions as described in section III-A. With these cost functions, classical motion planners such as Covariant Hamiltonian Optimization [17] are able to generate trajectories that maximize $p(g^*|\xi_{s_0->s_t})$, the objective for legible motion. In our method, we directly learn this distribution from the training data. The authors of [18] use an actor critic approach to train a legible policy in an online setting. In this paper we are specifically interested in learning an end-to-end *offline* policy for legible motion.

B. Offline Imitation Learning

Imitation Learning (IL) is a paradigm for learning from datasets of state action pairs which have been collected by expert demonstrators. Empirically, studies have shown that IL can achieve state-of-the-art performance across a variety of tasks even with sub-optimal data [19], [8]. In the field of HRI, learning from demonstrations is an important tool as it allows for non-expert programming of desired behaviors through kinesthetic teaching, teleoperation, or passive observation [20]. Another important factor is the safety afforded by offline learning. Deploying agents that learn through environment interaction is dangerous because actions with low reward (such as hitting a human) may be taken in the process of exploration. This danger is mitigated with offline IL as no environment exploration is necessary. From a deep learning perspective, the benefits to learning from offline datasets are scalability, portability, and reproducability. Compiling larger datasets is routinely used to dramatically improve deep vision and language models [21], [22]. There are ongoing efforts to collect similar data for robotics [7].

In this paper, we are concerned with training an agent to generate legible motion from multi-modal and multi-task datasets. Multi-modal distributions don't have a singular deterministic action output; rather, there can be multiple plausible actions from any given state. Algorithms such as Implicit Behavioral Cloning [8], Conditional Behavioral Transformers (C-BeT) [9], and Diffusion Policy [10] are all capable of learning effective policies from such distributions. In particular, algorithms based on Denoising Diffusion Probabilistic Models (DDPMs) [23], such as Diffusion Policy [10], have emerged as state of the art deep generative models for offline learning. A unique aspect of DDPMs is their ability to be guided through classifier free guidance, which allows for controllable generation at evaluation time [13]. Diffusion model guidance has proven useful for a range of tasks including offline reinforcement learning [24] and image generation [13]. With Legibility Diffuser, we show that guided generation from diffusion models can produce intent-expressive motion.

III. PRELIMINARIES

A. Equations for Legible Motion

Mathematically, a legible trajectory ξ from start state s_0 to goal state g^* optimizes the following equation [1]:

$$legibility(\xi) = \frac{\int p(g^*|\xi_{s_0 \to s_t}) f(t) dt}{\int f(t) dt} \tag{1}$$

Here $f(t)$ is a function of time that puts higher weight on earlier parts of the trajectory. Typically, $p(g|\xi_{s_0 \to s_t})$ is estimated using a cost function ζ that models what the observer expects the robot to do:

$$p(g|\xi_{s_0 \to s_t}) \propto \frac{exp(-\zeta[\xi_{s_0 \to s_t}] - v_g(s_t))}{exp(-v_g(s_0))} p(g) \tag{2}$$

where $v_y(x)$ is the lowest cost path from x to y. ζ is estimated and verified through user studies. To maximize $p(g^*|\xi_{s_0 \to s_t})$, one must minimize $p(g \neq g^*|\xi_{s_0 \to s_t})$, i.e., the probability of going to opposing goals. This is done by following an ongoing path $\xi_{s_0 \to s_t}$ such that $v_{g \neq g^*}(s_t) \gg v_{g^*}(s_t)$. For pick and place tasks, experiments [25] show:

$$\zeta[\xi] = \frac{1}{2} \int \xi'(t)^2 dt \tag{3}$$

A trajectory is legible if the cost of reaching an opposing goal is high while the cost of reaching the target goal is low. From this equation, it is clear that longer, slower paths have higher costs. For a pick-and-place task, straight line paths that move quickly toward an object will have a low cost.

While these methods are useful for estimating $p(g^*|\xi_{s_0 \to s_t})$, deep learning models should be able to directly learn this distribution from a training dataset. In this paper, we present a generative learning framework where

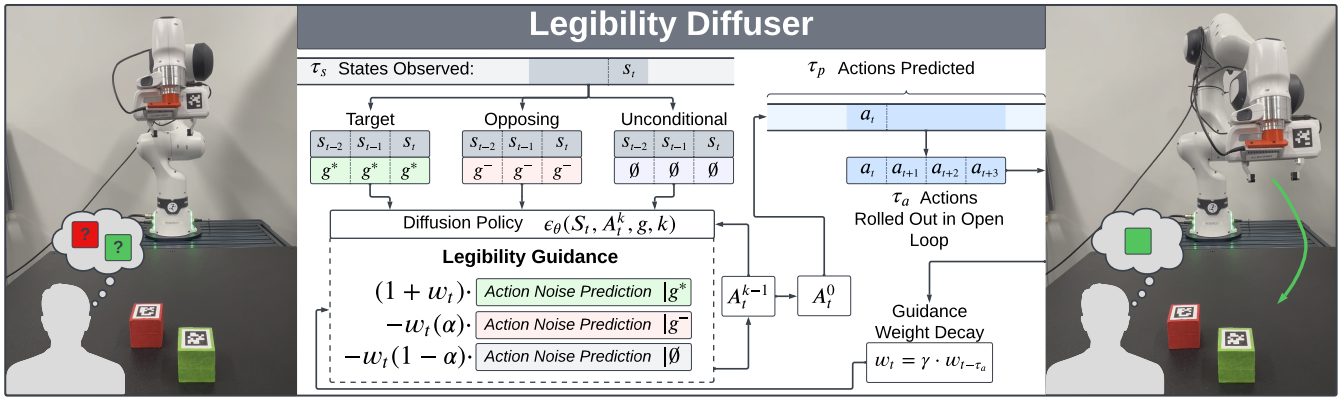

Fig. 2. **Legibility Diffuser:** This diagram shows the evaluation process for Legibility Diffuser. At each step, the model takes as input a sequence of τ_s states as well as a target goal g^* and an opposing goal g^-. While generating a sequence of actions, we use diffusion model guidance to ensure that we imitate the most legible actions from the training data. This is controlled by the guidance weight w_t and the ratio term α which determines the portion of negative guidance that goes to g^- versus $\emptyset$. Once the actions are generated, we carry out τ_a of the τ_p predicted actions in open loop. The guidance weight is decayed by γ before generating the next set of actions.

the learned generative policy is guided to directly generate actions which maximize $p(g^*|\xi_{s_0 \to s_t})$.

B. Sequential Decision Making

We view robot action generation as a sequential decision making problem and model it as a discrete-time infinite-horizon Markov Decision Process (MDP), $\mathcal{M} = (\mathcal{S}, \mathcal{A}, \mathcal{T}, \rho_0)$, where $\mathcal{S}$ is the state space, $\mathcal{A}$ is the action space, $\mathcal{T}(\cdot|s,a)$ is the state transition distribution, and $\rho_0(\cdot)$ is the initial state distribution. At every step, an agent observes a state s_t and queries a policy π to choose an action $a_t = \pi(s_t)$. The agent performs the action and observes the next state $s_{t+1} \sim \mathcal{T}(\cdot|s_t, a_t)$.

We augment this MDP with a set of absorbing goal states $G \subset \mathcal{S}$, where $g \in G$ corresponds to a specific state of the world in which the task is considered to be solved. Every pair (s_0, G) of an initial state $s_0 \sim \rho_0(\cdot)$ and goals for a task G corresponds to a new task instance. For the purposes of legibility, we also define a target goal $g^* \in G$ and an opposing goal $g^- \in G$. We want our motion to g^* to be distinct from motion to g^- such that $p(g^*|s_t, a_t) > p(g^-|s_t, a_t)$. We assume access to a dataset of N task demonstrations $D = \{\xi_i\}_{i=1}^N$ where each demonstration is a trajectory $\xi_i = (s_{i0}, a_{i0}, s_{i1}, a_{i1}, \ldots, s_{iT})$ that begins in a start state $s_{i0} \sim \rho_0(\cdot)$ and terminates in a goal state $s_{iT} = g_i$.

C. Conditional Denoising Diffusion Probabilistic Models

Conditional DDPMs [23] aim to estimate an unknown conditional distribution $q(\mathbf{x}_0|\mathbf{c})$ using a parameterized model $\pi_\theta(\mathbf{x}_0, \mathbf{c})$ based on sampled data $\mathbf{x}_0$ drawn jointly with conditioning information $\mathbf{c}$. The process consists of a *forward noising process* and a *reverse denoising process*. The forward process injects Gaussian noise into samples, producing noised distributions $q_t(\mathbf{x}_t|\mathbf{c})$. The distribution of $\mathbf{x}_t$ based on $\mathbf{x}_0$ is given as $q_{0t}(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \mathbf{x}_0, \sigma_t^2 \mathbf{I})$. During the reverse denoising process, a prediction network ϵ_θ estimates the noise added at time t: $\epsilon_\theta(\mathbf{x}_t, t, \mathbf{c}) \approx \nabla_{\mathbf{x}} \log q_t(\mathbf{x}_t|\mathbf{c})$. This allows the uncorrupted data $\mathbf{x}_0$ to be recovered using a stochastic differential equation originally laid out in [23].

$$\mathbf{x}_{t-1} = \phi_t(\mathbf{x}_t - \psi_t \epsilon_\theta(\mathbf{x}_t, t, \mathbf{c}) + \mathcal{N}(0, \sigma_t^2 I)) \tag{4}$$

Here ϕ_t, ψ_t, and σ_t are hyperparameters of the noise scheduling process that can be tuned. In classifier free guidance [13], the noise score $\bar{\epsilon_\theta}(\mathbf{x}_t, t, \mathbf{c})$ is calculated by combining the noise scores from a conditional estimate $\epsilon_\theta(\mathbf{x}_t, t, \mathbf{c})$ and an unconditional estimate $\epsilon_\theta(\mathbf{x}_t, t, \emptyset)$:

$$\bar{\epsilon_\theta}(\mathbf{x}_t, t, \mathbf{c}) = (1+w)\epsilon_\theta(\mathbf{x}_t, t, \mathbf{c}) - w\epsilon_\theta(\mathbf{x}_t, t, \emptyset) \tag{5}$$

The unconditional score estimation $\epsilon_\theta(\mathbf{x}_t, t, \emptyset)$ is trained at the same time as the conditional score estimation $\epsilon_\theta(\mathbf{x}_t, t, \mathbf{c})$ by setting class conditioning information $\mathbf{c}$ to a null token $\emptyset$ with probability p_{uncond}. Guidance weight w has the effect of up-weighting the probability of data for which a classifier $p_\theta(\mathbf{c}|\mathbf{x}_t)$ would assign high likelihood to the correct label. This formulation has the benefit of not requiring a classifier trained on partially corrupted data $\mathbf{x}_t$.

IV. LEGIBILITY DIFFUSER

A. Overview

The key contribution of our paper is framing legible motion generation as the output of guided conditional diffusion policies. As described in section III-C, a conditional DDPM π_θ estimates an unknown conditional distribution $q(\mathbf{x}_0|\mathbf{c})$ by drawing samples $\mathbf{x}_0$ jointly with class conditioning information $\mathbf{c}$. These models can be guided to generate outputs such that a classifier $p_\theta(\mathbf{c}|\mathbf{x}_0)$ trained on the same samples as π_θ would assign a high likelihood to the correct class label [13]. In the setting of robotic policy generation, we train a generative model to estimate $q(a_t|s_t, g^*)$, the distribution of actions a_t given the current state s_t and a target goal g^*. Using diffusion guidance, we can generate outputs such that a classifier $p_\theta(g|a_t, s_t)$ would assign a high probability to g^*. We notice that this resembles the objective of legible motion generation, i.e. maximizing $p(g^*|\xi_{s_0 \to s_t})$ (Eq. 2) where $\xi_{s_0 \to s_t}$ is the ongoing trajectory. In this formulation, (a_t, s_t) can be thought of as a small section (the most recent section) of $\xi_{s_0 \to s_t}$. If there is a human observing the

robot, then this collaborator is the classifier $p_\theta(g|a_t, s_t)$ that assigns a high probability to g^*. Below we detail how our method, Legibility Diffuser, is able to leverage this property of diffusion model guidance to generate legible robot motion.

B. Conditional Generative Policy Formulation

Legibility Diffuser (Fig. 2) is a DDPM that serves as a visuomotor robot policy. Our work builds upon recent advancements in DDPM-based policies, most notably Diffusion Policy [10]. At each time step, we generate actions $A_t = a_{t:t+\tau_p}$ for an agent given initial states $S_t = s_{t:t+\tau_s}$ and a goal state g. This is done by training a UNet-based diffusion model policy $\pi_\phi(A_t|S_t, g)$. Once the actions are generated, the agent carries out $\tau_a < \tau_p$ steps in open loop. Training samples are drawn from the demonstration dataset by sampling sub-sequences of length τ_p that include actions, states, and the goal state. We define the goal state g as the final state in the demonstration. We construct our conditioning term by concatenating S_t with τ_s repetitions of g. In the style of classifier free guidance [13], we zero out g to train $\pi_\phi(A_t|S_t, \emptyset)$ with probability p_{uncond}. Our training procedure follow the standard formulation for DDMP training as laid out in III-C. Generating legible motion does not require any additional steps during training. At evaluation, we require access to a target goal and opposing goal(s) in order to guide π_ϕ to generate legible actions.

C. Legibility Guidance

We guide our diffusion model policy π_ϕ towards the most legible action sequence A_t at each timestep t. To do this, we require access to a target goal g^* that we hope our agent reaches. We also require an opposing goal g^- that we do not want our agent to reach. For our motion to be legible, at every timestep t an observer should be able to predict that the agent is going to g^* and not g^-. In other words, $p(g^*|S_t, A_t) > p(g^-|S_t, A_t)$. Again, because we are only looking at (S_t, A_t), the most recent segment of the entire trajectory $\xi_{s_0 \to s_t}$, this is a slight simplification of the true legibility objective (Eq. 1) which maximizes $p(g^*|\xi_{s_0 \to s_t})$. We achieve this maximization through diffusion model guidance.

In previous works, "A and not B" diffusion model compositions have been used to remove undesirable outputs during image [26], language [27], and robotic action generation [24]. Inspired by this, we generate legible motion through "g^* and not g^-" compositions. At each denoising step k, we calculate a noise score conditioned on the target goal $\epsilon_\theta(g^*) := \epsilon_\theta(A_t^k, S_t, g^*, k)$, the opposing goal $\epsilon_\theta(g^-) := \epsilon_\theta(A_t^k, S_t, g^-, k)$, and a null token $\epsilon_\theta(\emptyset) := \epsilon_\theta(A_t^k, S_t, \emptyset, k)$. We combine these scores in a manner similar to classifier free guidance (Eq. 5) to get a final noise score as follows:

$$\bar{\epsilon}_\theta(g^*) = (1+w_t)\epsilon_\theta(g^*) - \alpha w_t \epsilon_\theta(g^*) - (1-\alpha) w_t \epsilon_\theta(\emptyset) \quad (6)$$

Here, α_t and w_t are hyperparameters tuned via a randomized grid search. The noise score is used to recover an uncorrupted action sequence A_t^0 following the standard stochastic process for DDPMs (Eq. 4). By increasing w_t, we guide π_ϕ to generate actions that are distinct from the outputs of both an unconditional model and a model conditioned on opposing goal g^-. α is a ratio term that determines the portion of guidance that goes to $\epsilon_\theta(g^-)$ vs. $\epsilon_\theta(\emptyset)$. We expect $\alpha \approx 1$ as guiding away from $\epsilon_\theta(g^-)$ should directly lead to actions where an observer predicts $p(g^*|S_t, A_t) > p(g^-|S_t, A_t)$ (sec. IV-A).

D. Time-varying Legibility Guidance Decay

Inspired by techniques for legible motion generation in HRI, we introduce a decaying term γ to our guidance weights. This causes legibility guidance to be strongest at the beginning of the trajectory. Various HRI papers have shown that maximizing legibility is most important at the beginning of a trajectory [1], [4] as we want observers to *quickly* infer the goal state. The guidance weight only decays after τ_a action steps, it is constant for every denoising step while generating the open loop action sequence of length τ_p:

$$w_t = \gamma \cdot w_{t-\tau_a} \quad (7)$$

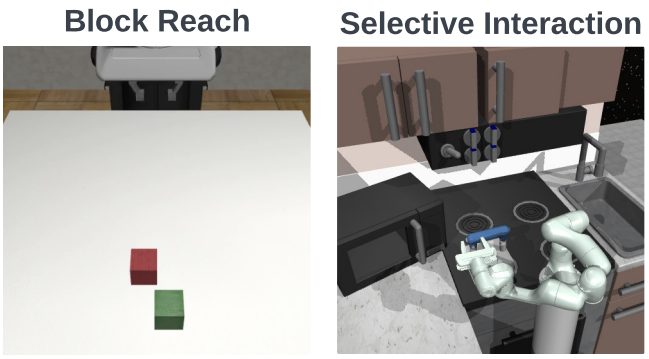

Fig. 3. **Environment Visualization:** Legibility Diffuser is evaluated on two tasks in two environments. We collect our own demonstration dataset for block reach and we use the original Franka Kitchen dataset [28] for selective interaction. Real world block reach visualization in Fig 1.

V. METHODS

A. Motivation

Through our experiments, we aim to show that Legibility Diffuser can clone the most legible mode in a demonstration dataset without sacrificing success rate. We evaluate our method on four tasks across two environments in simulation as well as one real world experiment. Legibility is measured autonomously through hand designed score functions.

B. Tasks and Environments

For each task, we define a target goal state $g*$ for the agent to reach and an opposing goal state g^- for the agent to avoid. We assume access to low dimensional states and proprioception. All data is collected through human teleoperation. Data is multi-modal in the sense that demonstrations show multiple ways of reaching a goal.

Block Reach: This task is based on a frequently used legibility experiment [1]. Two blocks are placed next to each-other on a table, a robot arm is positioned on the other end of the table. The robot reaches for one of the blocks and lifts it. Assume there is an observer. If the agent moves directly towards the blocks, it is unclear if the target is the left block or the right block. If the agent takes a trajectory that swings wide to one side, then g^* can be easily predicted. These wide trajectories are the legible trajectories.

Real: We run a real world experiment on a Franka Emika Panda Robot using an OSC pose controller. We use April-Tags [29] to predict low dimensional states, allowing for sim-to-real transfer. 100 demonstrations are used to train the agent and all roll-outs are run in open loop.

Robosuite: We use Robosuite [30] for our block reach simulation and for sim-to-real transfer. Our block reach simulation agent is trained on 200 demonstrations.

Selective Interaction (Franka Kitchen): In this task, there are m objects in a scene $o_1, o_2, ..., o_m$ and each object has a corresponding goal position. The robot must move $n < m$ objects to their goal positions (within tolerance ϵ_{tol}). A goal g can be described by this set of n objects. g^* and g^- have $n-1$ objects in common; there is only one distinguishing object. We define o^* as the distinguishing object for g^* and o^- as the distinguishing object for g^-. For example, if $g^* = \{o_1, o_2, o_3\}$ and $g^- = \{o_1, o_2, o_4\}$, $o^* = o_3$ and $o^- = o_4$. A legible agent *should* interact with o^* early in the trajectory so that an observe can quickly predict the target goal. A legible agent *should not* interact with o^- as this would confuse the observer, resulting in an incorrect goal prediction. This task evaluates long horizon legibility.

We adapt the Franka Kitchen environment and dataset [28] for selective interaction. In Franka Kitchen there are 7 objects ($m = 7$) and we set $n = 5$ for the goal states. The demonstrations show the robot interacting with four objects. The four objects differ across demonstrations but always follow a fixed interaction order ($o_1 \rightarrow o_2 \rightarrow ... \rightarrow o_7$). We define three experiments with the level of difficulty based on where o^* and o^- occur in this interaction order:

Kitchen Easy: $o^* = o_1$ and $o^- = o_2$. All the decisions that impact legibility have a short horizon.

Kitchen Medium: $o^* = o_3$ and $o^- = o_4$. Critical interactions are in the middle of the interaction sequence.

Kitchen Hard: $o^* = o_6$ and $o^- = o_5$. All decisions impacting legibility have a long horizon.

C. Metrics

For each task we define a legibility score function which captures the notion that an observer should predict $p(g^*) > p(g^-)$. We use this score function for automated evaluation of legibility. This allows us to benchmark a wider ranger of algorithms and perform larger scale ablations. We also evaluate task success rate.

Legibility Score - Block Reach: We use a distance-based heuristic for this task. Many experiments have shown that for pick and place tasks, legible trajectories maximize the distance to the opposing goal [1]. We draw inspiration from these papers and calculate legibility as follows:

$$L(\xi_{s_0 \rightarrow g^*}) = \sum_{s_t \in \xi_{s \rightarrow g^*}} \frac{||g^- - s_t||_2}{t} \tag{8}$$

The legibility values reported are normalized based on the minimum and maximum values in the training dataset.

Legibility Score - Selective Interaction: We record t_{o^*} and t_{o^-}, the time-step at which the agent interacts with o^* and o^- respectively. We define the legibility score as:

$$L(\xi_{s_0 \rightarrow g^*}) = \frac{t_{max} - t_{o^*}}{t_{max}} \cdot \mathbb{1}_{o^* < o^-} \tag{9}$$

Here t_{max} is the maximum time-step in the training dataset and $t_{o^*} \rightarrow t_{max}$ if the agent never interacts with o^*. Failure to interact with o^* or engaging with o^- before o^* both result in a legibility score of 0. For example, in *Kitchen Hard* $o^* = o_6$ and $o^- = o_5$. So roll-outs $o_1 \rightarrow o_2 \rightarrow o_5 \rightarrow o_6$ and $o_1 \rightarrow o_2 \rightarrow o_3$ both get a legibility score of 0.

Success: For block reach we report roll-out success rate. A roll-out is successful if the conditioned goal state is reached before t_{max}. For selective interaction, we report the average number of objects in g^* moved to their goal position.

D. Baselines

Our baselines evaluate if Legibility Diffuser is more capable of generating legible motion than other offline IL algorithms. We train on multi-modal, multi-task data and compare against algorithms that perform well on such data.

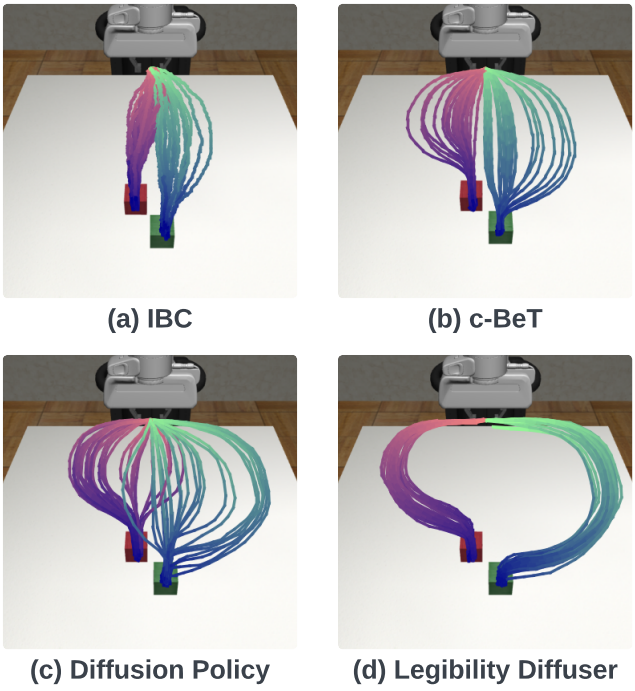

Fig. 4. **Block Reach Rollouts**: 50 rollouts of Legibiltiy Diffuser and baselines in simulation. C-Bet and Diffusion Policy capture the entire training data distribution, IBC collapses on an arbitrary mode, Legibility Diffuser collapses on the most legible mode.

	Block Reach		Selective Interaction		
	Real	Sim	Easy	Medium	Hard
IBC [8]	-	.58 ± .13	0.0 ± .00	0.0 ± .00	0.0 ± .00
c-BeT [9]	-	.60 ± .21	.57 ± .38	.02 ± .08	.11 ± .14
Diffusion Policy [10]	.42 ± .15	.77 ± .29	.67 ± .34	.39 ± .15	.22 ± .12
Legibility Diffuser (Ours)	**.68 ± .18**	**1.14 ± .15**	**.86 ± .03**	**.55 ± .09**	**.26 ± .08**
Oracle	1.0 ± 0.0	1.0 ± 0.0	.83 ± .01	.78 ± .05	.73 ± .01
Dataset	.45 ± .23	.45 ± .23	.60 ± .37	.42 ± .26	.26 ± .11

TABLE I

LEGIBILITY SCORES FOR OUR MODEL AND BASELINES AVERAGED OVER 50 SEEDS. THESE VALUES ARE CALCULATED USING THE LEGIBILITY SCORE FUNCTIONS (EQ. 8, 9). BLOCK REACH REAL IS AVERAGED OVER 3 SEEDS AND 30 DEMONSTRATIONS.

	Block Reach		Selective Interaction		
	Real	Sim	Easy	Medium	Hard
IBC [8]	-	.92	.02 ± .14	.04 ± .20	0.0 ± 0.0
c-BeT [9]	-	.98	1.1 ± .83	.44 ± 1.02	2.22 ± 1.77
Diffusion Policy [10]	**.97**	**1**	3.32 ± .97	**3.54 ± .61**	**3.86 ± .45**
Legibility Diffuser (Ours)	**.97**	**1**	**4.00 ± .35**	**3.48 ± 0.78**	**3.90 ± .36**

TABLE II

SUCCESS RATE OF OUR MODEL AND BASELINES AVERAGED OVER 50 SEEDS. FOR SELECTIVE INTERACTION WE REPORT THE AVERAGE NUMBER OF OBJECT INTERACTIONS. BLOCK REACH REAL IS AVERAGED OVER 3 SEEDS AND 30 DEMONSTRATIONS.

Conditional Behavioral Transformers (c-BeT) [9]: c-BeT is a behavioral cloning algorithm with a transformer backbone. This algorithm discretizes the action space and generates roll-outs using action chunking. It is known to learn effective conditional policies from multi-task, multi-modal play data.

Implicit Behavioral Cloning (IBC) [8]: IBC is an energy based behavioral cloning algorithm. We condition this model in the same manner as Legibility Diffuser, concatenating the target goal state to the current state

Diffusion Policy [10]: Diffusion policy is the DDPM that Legibility Diffuser is built on. We implement and evaluate a goal conditioned version of diffusion policy, isolating the effect of our contributions. This is equivalent to Legibility Diffuser with guidance weight (w) set to zero.

Classical Legibility (Oracle): Classical legibility methods require oracle access to the underlying cost function that captures an observer's expectations (Section III-A). Therefore, this baseline is assumed to perform optimally. For the block reach task, this is an agent that imitates the most legible trajectory in the training dataset. For selective interaction, this agent's first interaction is with o^*.

Training Dataset (Dataset): We report the legibility statistics of our training dataset.

E. Ablation Study

We conduct a comprehensive hyperparameter ablation on two tasks in simulation, block reach and selective interaction medium. We perform an ablation over α (Eq. 6) and γ (Eq. 7) as we vary guidance weight w. When we ablate over α, γ is set to the tuned value from training (and visa versa).

	Block Reach		Selective Interaction		
	Real	Sim	Easy	Medium	Hard
w	5.0	7.5	5.0	5.0	1.0
α	0.9	1.0	0.9	0.9	0.9
γ	0.5	0.5	0.75	0.95	1.0

TABLE III

TUNED HYPERPARAMETERS (EQ. 6, 7) FOR LEGIBILITY DIFFUSER

VI. RESULTS

Legibility Diffuser produces legible robot motion. As shown in Table I, our method generates trajectories that are more legible than any of the imitation learning baselines. Results from t-test analysis reveal that for all tasks, the differences between our method and the baselines are statistically significant ($p < .05$). Additionally, across all tasks we meet or exceed the average legibility in the training dataset. For *block reach - sim* and *selective interaction - easy*, Legibility Diffuser even achieves oracle performance and matches the expected legibility of classical methods. However, our method is certainly still constrained by the training data. This is best seen with *selective interaction - medium and hard* where our agent is unable to go directly to o^*. The demonstration dataset lacks these legible modes, and generalization of generative models is an open area of research. A dataset that is more multi-modal than Franka Kitchen is likely better suited for the selective interaction task. Overall, our method is able to imitate legible behaviors from a multi-modal dataset without any access to the cost functions and classical motion planners typically required for legible motion generation.

Optimizing for legibility does not sacrifice success rate. From Table I we see that our method is able to maintain state

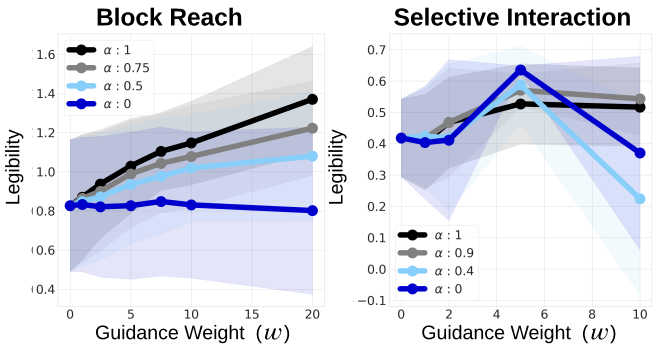

Fig. 5. **Alpha Ablation - Legibility**: We ablate over α while varying guidance weight w (Eq. 6). $\alpha \approx 1$ gives us a high legibility on both tasks across all of the guidance weights.

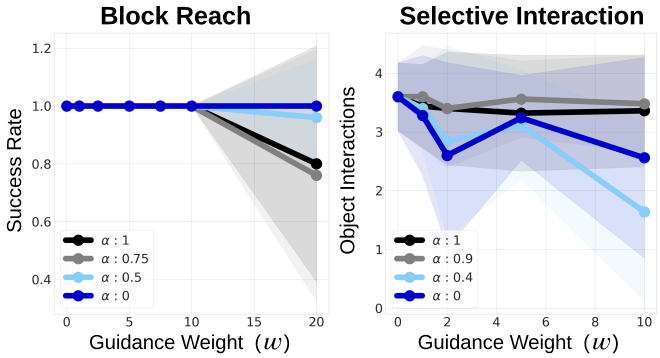

Fig. 6. **Alpha Ablation - Success**: We ablate over α while varying guidance weight w (Eq. 6). For most guidance weights, success rate is not particularly sensitive to α.

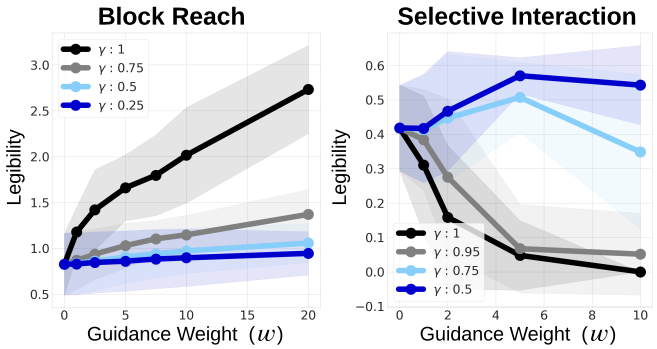

Fig. 7. **Gamma Ablation - Legibility**: We ablate over γ while varying guidance weight w (Eq. 7). $\gamma \to 1$ leads to higher legibility as long as success rate is maintained (Fig. 8).

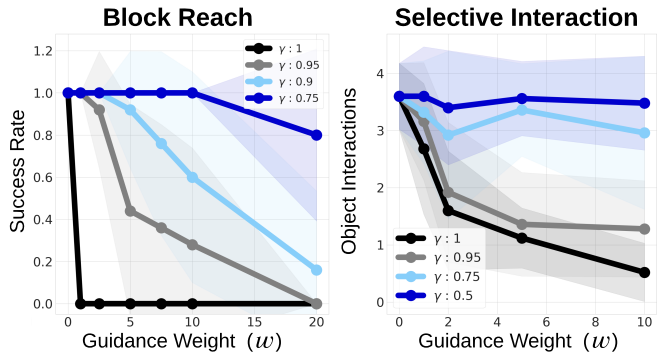

Fig. 8. **Gamma Ablation - Success**: We ablate over γ while varying guidance weight w (Eq. 7). Across all guidance weights we see that success rate decreases as $\gamma \to 1$.

of the art success rates across all tasks while maximizing legibility. For *selective interaction - easy* we actually see that Legibility Diffuser achieves the highest success rate. We hypothesize that this is because classifier free guidance is also useful for constraint satisfaction [24]. Without this stronger conditioning, the model will occasionally interact with an object that is not in the goal state. IBC and c-BeT both struggle to complete the longer horizon *selective interaction* tasks. This is exasperated by the fact that the conditioned goal has five objects, which is slightly out of distribution of the training data with four objects.

Increasing guidance weight increases legibility: In both our α (Fig. 5) and γ (Fig. 7) ablations, we notice a positive correlation between legibility and guidance weight. This suggests that legibility guidance is successfully pushing the agent to the most legible modes in the data. The trend only breaks down at high guidance weights when we see success rates decrease as well (Fig. 6, 8). Specifically, this happens for the selective interaction task which requires successful object interactions to evaluate legibility (Eq. 9).

Time-varying guidance weight is a critical component of Legibility Diffuser. Our ablation study over γ shows that guidance weight decay has a large effect on the performance of the model. For both block reach and selective interaction, we notice that increasing the guidance weight causes the success rate to decrease. The effect is most pronounced when $\gamma = 1$ (Fig. 8). Empirically, we find that high guidance weights throw the agent out of distribution, and it is often unable to recover. By decreasing the guidance weight over the course of the trajectory ($\gamma < 1$), we are able to maintain a high success rate while having a large initial guidance weight (Fig. 8). These high guidance weights are critical for legibility, so proper tuning of γ allows us to get good performance on both metrics at the same time. We find that the appropriate decay rate, γ, is relatively task specific. Generally, for longer horizon tasks, $\gamma \to 1$.

Negative guidance on opposing goal g^- is helpful for Legibility Diffuser: When ablating over α (Fig 5, 6), we observe less consistent trends compared to our ablation over γ. However, we achieve good performance for both tasks and metrics when $\alpha \approx 1$. This means that the majority of our guidance weight, w, is directed towards guiding away from g^- (Eq. 6).

VII. DISCUSSION AND FUTURE WORK

With Legibility Diffuser, we show that legible robot motion can be generated directly from a dataset of multi-modal, multi-task human demonstrations. This data driven approach does not require access to the hand designed cost functions or classical motion planners that are typically needed for legible

motion generation. We evaluate our model on tasks where an observer's cost function is easy to estimate, but Legibility Diffuser can be applied to any task or mulit-modal dataset.

We show that the objective for guided diffusion models matches the objective for legible motion generation. By placing a large negative guidance weight on $\epsilon_\theta(g^-)$ (the noise score of a model conditioned on opposing goal g^-), we are able to generate legible robot motion. Decaying this guidance weight over time allows us to maintain competitive success rates. Without weight decay ($\gamma < 1$), high guidance weights lead to a trade-off between legibility and success rate. This suggests that standard classifier free guidance, which does not have weight decay, may not be ideal for imitation learning. Further investigation of this phenomenon on a larger range of datasets is needed before concrete conclusions can be drawn. Our experiments focus on legibility, so we make use of domain knowledge and have high guidance only at the beginning of the trajectory, where legibility is most important [1], [4]. For other tasks and scenarios, learning a state conditioned guidance weight may be more appropriate.

Our experiments show that the effectiveness of Legibility Diffuser depends on the diversity of the demonstration dataset. This is evident by the limited legibility in the Franka Kitchen environment. In future work we plan to train on larger datasets with greater multi-modality. Additionally, we hope to evaluate the legibility of our method through an in-depth user study. The key takeaway from this paper is that given a multi-task, multi-modal dataset, guided diffusion models can produce actions that maximizes the same objective as legible robot motion. We hope Legibility Diffuser is useful for future deep learning approaches to HRI.

REFERENCES

[1] A. D. Dragan, K. C. Lee, and S. S. Srinivasa, "Legibility and predictability of robot motion," in *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 301–308, IEEE, 2013.

[2] C. Lichtenthäler, T. Lorenz, and A. Kirsch, "Towards a legibility metric: How to measure the perceived value of a robot," in *International Conference on Social Robotics, ICSR 2011*, 2011.

[3] C. Breazeal, C. D. Kidd, A. L. Thomaz, G. Hoffman, and M. Berlin, "Effects of nonverbal communication on efficiency and robustness in human-robot teamwork," in *2005 IEEE/RSJ international conference on intelligent robots and systems*, pp. 708–713, IEEE, 2005.

[4] A. D. Dragan, S. Bauman, J. Forlizzi, and S. S. Srinivasa, "Effects of robot motion on human-robot collaboration," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, pp. 51–58, 2015.

[5] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M.-A. Lachaux, T. Lacroix, B. Rozière, N. Goyal, E. Hambro, F. Azhar, *et al.*, "Llama: Open and efficient foundation language models," *arXiv preprint arXiv:2302.13971*, 2023.

[6] P. Dhariwal and A. Nichol, "Diffusion models beat gans on image synthesis," *Advances in Neural Information Processing Systems*, vol. 34, pp. 8780–8794, 2021.

[7] A. Padalkar, A. Pooley, A. Jain, A. Bewley, A. Herzog, A. Irpan, A. Khazatsky, A. Rai, A. Singh, A. Brohan, *et al.*, "Open x-embodiment: Robotic learning datasets and rt-x models," *arXiv preprint arXiv:2310.08864*, 2023.

[8] P. Florence, C. Lynch, A. Zeng, O. A. Ramirez, A. Wahid, L. Downs, A. Wong, J. Lee, I. Mordatch, and J. Tompson, "Implicit behavioral cloning," in *Conference on Robot Learning*, pp. 158–168, PMLR, 2022.

[9] Z. J. Cui, Y. Wang, N. M. M. Shafiullah, and L. Pinto, "From play to policy: Conditional behavior generation from uncurated robot data," *arXiv preprint arXiv:2210.10047*, 2022.

[10] C. Chi, S. Feng, Y. Du, Z. Xu, E. Cousineau, B. Burchfiel, and S. Song, "Diffusion policy: Visuomotor policy learning via action diffusion," *arXiv preprint arXiv:2303.04137*, 2023.

[11] K. Grauman, A. Westbury, E. Byrne, Z. Chavis, A. Furnari, R. Girdhar, J. Hamburger, H. Jiang, M. Liu, X. Liu, *et al.*, "Ego4d: Around the world in 3,000 hours of egocentric video," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 18995–19012, 2022.

[12] C. Lynch, M. Khansari, T. Xiao, V. Kumar, J. Tompson, S. Levine, and P. Sermanet, "Learning latent plans from play," in *Conference on robot learning*, pp. 1113–1132, PMLR, 2020.

[13] J. Ho and T. Salimans, "Classifier-free diffusion guidance," *arXiv preprint arXiv:2207.12598*, 2022.

[14] M. Tomasello, M. Carpenter, J. Call, T. Behne, and H. Moll, "Understanding and sharing intentions: The origins of cultural cognition," *Behavioral and brain sciences*, vol. 28, no. 5, pp. 675–691, 2005.

[15] G. Hoffman and G. Weinberg, "Shimon: an interactive improvisational robotic marimba player," in *CHI'10 Extended Abstracts on Human Factors in Computing Systems*, pp. 3097–3102, 2010.

[16] L. Takayama, D. Dooley, and W. Ju, "Expressing thought: improving robot readability with animation principles," in *Proceedings of the 6th international conference on Human-robot interaction*, pp. 69–76, 2011.

[17] M. Zucker, N. Ratliff, A. D. Dragan, M. Pivtoraiko, M. Klingensmith, C. M. Dellin, J. A. Bagnell, and S. S. Srinivasa, "Chomp: Covariant hamiltonian optimization for motion planning," *The International Journal of Robotics Research*, vol. 32, no. 9-10, pp. 1164–1193, 2013.

[18] X. Zhao, T. Fan, D. Wang, Z. Hu, T. Han, and J. Pan, "An actor-critic approach for legible robot motion planner," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5949–5955, IEEE, 2020.

[19] A. Mandlekar, D. Xu, J. Wong, S. Nasiriany, C. Wang, R. Kulkarni, L. Fei-Fei, S. Savarese, Y. Zhu, and R. Martín-Martín, "What matters in learning from offline human demonstrations for robot manipulation," *arXiv preprint arXiv:2108.03298*, 2021.

[20] H. Ravichandar, A. S. Polydoros, S. Chernova, and A. Billard, "Recent advances in robot learning from demonstration," *Annual review of control, robotics, and autonomous systems*, vol. 3, pp. 297–330, 2020.

[21] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255, Ieee, 2009.

[22] L. Floridi and M. Chiriatti, "Gpt-3: Its nature, scope, limits, and consequences," *Minds and Machines*, vol. 30, pp. 681–694, 2020.

[23] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in neural information processing systems*, vol. 33, pp. 6840–6851, 2020.

[24] A. Ajay, Y. Du, A. Gupta, J. Tenenbaum, T. Jaakkola, and P. Agrawal, "Is conditional generative modeling all you need for decision-making?," *arXiv preprint arXiv:2211.15657*, 2022.

[25] A. Dragan and S. Srinivasa, "Familiarization to robot motion," in *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, pp. 366–373, 2014.

[26] P. Schramowski, M. Brack, B. Deiseroth, and K. Kersting, "Safe latent diffusion: Mitigating inappropriate degeneration in diffusion models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 22522–22531, 2023.

[27] T. Schick, S. Udupa, and H. Schütze, "Self-diagnosis and self-debiasing: A proposal for reducing corpus-based bias in nlp," *Transactions of the Association for Computational Linguistics*, vol. 9, pp. 1408–1424, 2021.

[28] A. Gupta, V. Kumar, C. Lynch, S. Levine, and K. Hausman, "Relay policy learning: Solving long-horizon tasks via imitation and reinforcement learning," *arXiv preprint arXiv:1910.11956*, 2019.

[29] E. Olson, "AprilTag: A robust and flexible visual fiducial system," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3400–3407, IEEE, May 2011.

[30] Y. Zhu, J. Wong, A. Mandlekar, R. Martín-Martín, A. Joshi, S. Nasiriany, and Y. Zhu, "robosuite: A modular simulation framework and benchmark for robot learning," *arXiv preprint arXiv:2009.12293*, 2020.