arXiv:2301.01576v1 [cs.RO] 4 Jan 2023

ROBOFRIEND：自适应讲故事机器人泰迪熊 – 技术报告

预印本

Ido Glanz*
Technion自主系统项目 以色列理工学院

Matan Weksler*
Technion自主系统项目 以色列理工学院

Erez Karpas
数据与决策科学学院 以色列理工学院 karpase@technion.ac.il

Tzipi Horowitz-Kraus
科学与技术教育学院 以色列理工学院
tzipi.kraus@ed.technion.ac.il

摘要

本文介绍了 Robofriend，这是一款为幼儿讲故事的机器人泰迪熊。Robofriend 利用强化学习调整自身行为，以保持儿童的注意力。

Keyw词机器人学 · 教育 · 强化学习

引言

早期发展阶段的语言接触对于促进与语言相关的脑网络形成至关重要（Kuhl [2004], Cardillo 和 Kuhl [2009], Moon 等， [2013]）。讲故事是一种语言接触形式，研究发现，它不仅与儿童在语言处理方面更高的参与度相关，也与其视觉化能力和认知能力相关（Hutton 等， [2015]）。有趣的是，有研究指出，与这些能力提升相关的并非讲故事本身，而是故事讲述过程中的互动；互动能够增强儿童的这些能力（Twait 等, [2019]）。一项近期研究表明，与以非互动方式参与讲故事活动（在屏幕上观看讲故事）的儿童相比，以互动方式参与活动的4至6岁儿童表现出更强的认知能力和语言能力（Twait 等， [2019]）。因此，人们提出了一个问题：互动式（对话式）讲故事带来的这种积极效果，究竟是源于人与人之间的互动，还是源于讲故事本身的互动性质？换句话说，在讲故事过程中使用互动机器人，能否取得与真人互动条件相似的效果?

为了研究这一问题，我们设计了 Robofriend（如图1所示）——一只为幼儿朗读故事的机器人泰迪熊。Robofriend 是在普通泰迪熊的肚子里装入一台平板电脑，并配备简易骨架、电机和传感器制成的，这些装置使它能够活动头部和手臂。Robofriend 可以为一小群儿童朗读故事，机器人的主要目标是吸引儿童，让他们专注于故事。因此，尽管我们设计 Robofriend 的主要动机是开展上述科学研究，它也可以作为托儿班教师的工具。Robofriend 可以为一组儿童朗读故事，同时教师则与班上的其他儿童互动。

Robofriend 能讲述的每个故事都分为多个预录视频片段。通常，每个片段会展示印刷书籍中某一页面的静态图像，同时由真人朗读页面上的文字。请注意，Robofriend 不会进行文本转语音，所有片段都是预先录制的。每个片段结束时，Robofriend 会从多个可用动作中选择一个来执行。可选动作包括提出一个关于故事的简单提问（每个故事片段都有一组预录问题，Robofriend 会从中随机选择一个

* 表示贡献相同

Item	Use	数量	价格
熊玩偶	-	1	US$30.00
显示器	播放视频	1	200.00 美元
摄像头	监控儿童	1	79.00 美元
Arduino + 接线套件	控制伺服电机	1	37.00 美元
伺服电机	移动头部	4	24.00 美元
扬声器	播放声音	1	20.00美元
总计：			462.00美元

表1：Robofriend 物料清单

），给予积极反馈（例如：“孩子们，真棒，我看得出你们在认真听”），或给予消极反馈（例如：“孩子们，你们在听故事吗？”）。
如前所述，Robofriend 的目标是保持儿童的注意力。优化事物的第一步是对其进行测量，或至少测量某种可作为其代理的指标。Robofriend 使用摄像头测量一些可作为参与度代理的指标。首先，Robofriend 使用计算机视觉检测儿童的人脸及其视线方向。需要说明的是，这些人脸都是匿名的——Robofriend 不会尝试以任何方式将人脸与身份关联起来。通过这些人脸检测结果，我们提取了几项测量值：有多少张人脸正看着机器人、他们对故事的专注程度（根据相对视线判断），以及人脸有多“躁动”（即“兴奋度”指标）。除了视觉属性外，我们还监测噪声水平及其瞬时变化（即其导数），作为辅助指标来反映儿童的状态。我们将这些指标汇总为每个摄像头视频帧对应的奖励信号，并在每个故事片段中进一步汇总，为每个片段生成一个状态和奖励。

在定义了奖励之后，我们现在可以尝试优化目标，即奖励的总和。当然，我们事先并不知道在每个故事片段之后应采取什么正确动作，也没有关于每个动作将如何影响儿童参与度的模型。因此，我们选择使用强化学习来控制 Robofriend 的动作。然而，由于涉及幼儿，我们不希望允许机器人探索那些我们知道对儿童无益的动作序列（例如始终使用消极反馈动作）。因此，我们采用了使用 LTL“约束机制”的方法，Giacomo 等人 [2020], 并手动编码 Robofriend 所允许的轨迹。

在本文的其余部分，我们将更详细地介绍 Robofriend 的设计。我们还将介绍机器人在当地托儿中心进行的初步评估。最后，我们将总结从该项目中汲取的一些经验，并讨论项目中出现的伦理考量。

Robofriend 设计

现在，我们将更详细地介绍 Robofriend 的设计，首先从机械构造开始。

Robofriend 机械构造

如前所述，Robofriend 是通过对一只身高 1 米的大型泰迪熊进行改造而构建的，使其能够移动头部和手臂、播放视频和声音，并注视听它讲故事的儿童。首先，我们在机器人内部装入一个铝制骨架，以支撑其他设备。为此，我们取出了大部分填充物，并暂时将头部拆开，以便将摄像头安装在泰迪熊的鼻子里，并为伺服电机安装固定架。随后，我们安装了一台 12.3 英寸的显示器，机器人会在上面显示故事片段；此外还安装了 4 个伺服电机，用于移动头部和手臂（每只手臂各一个，另外两个用于控制头部水平转动和俯仰）。如上所述，我们在泰迪熊的鼻子里安装了摄像头，用来监视儿童并测量奖励信号；最后，还连接了扬声器来播放声音。示意图见图2。

这些硬件由一台个人电脑控制，摄像头、显示器和扬声器均直接连接到该电脑。伺服电机由 Arduino Uno 控制（见图3），Arduino Uno 也连接到个人电脑。伺服电机的控制器在 Arduino 上作为独立进程运行，并遵循个人电脑发出的指令。

Robofriend 的物料清单（BOM）见表1，包含各商品链接的详细物料清单可在线查阅：https://shorturl.at/efBZ2。总体而言，制作 Robofriend 的总成本不到 500美元，因此相当容易负担。



图1：Robofriend 家庭测试

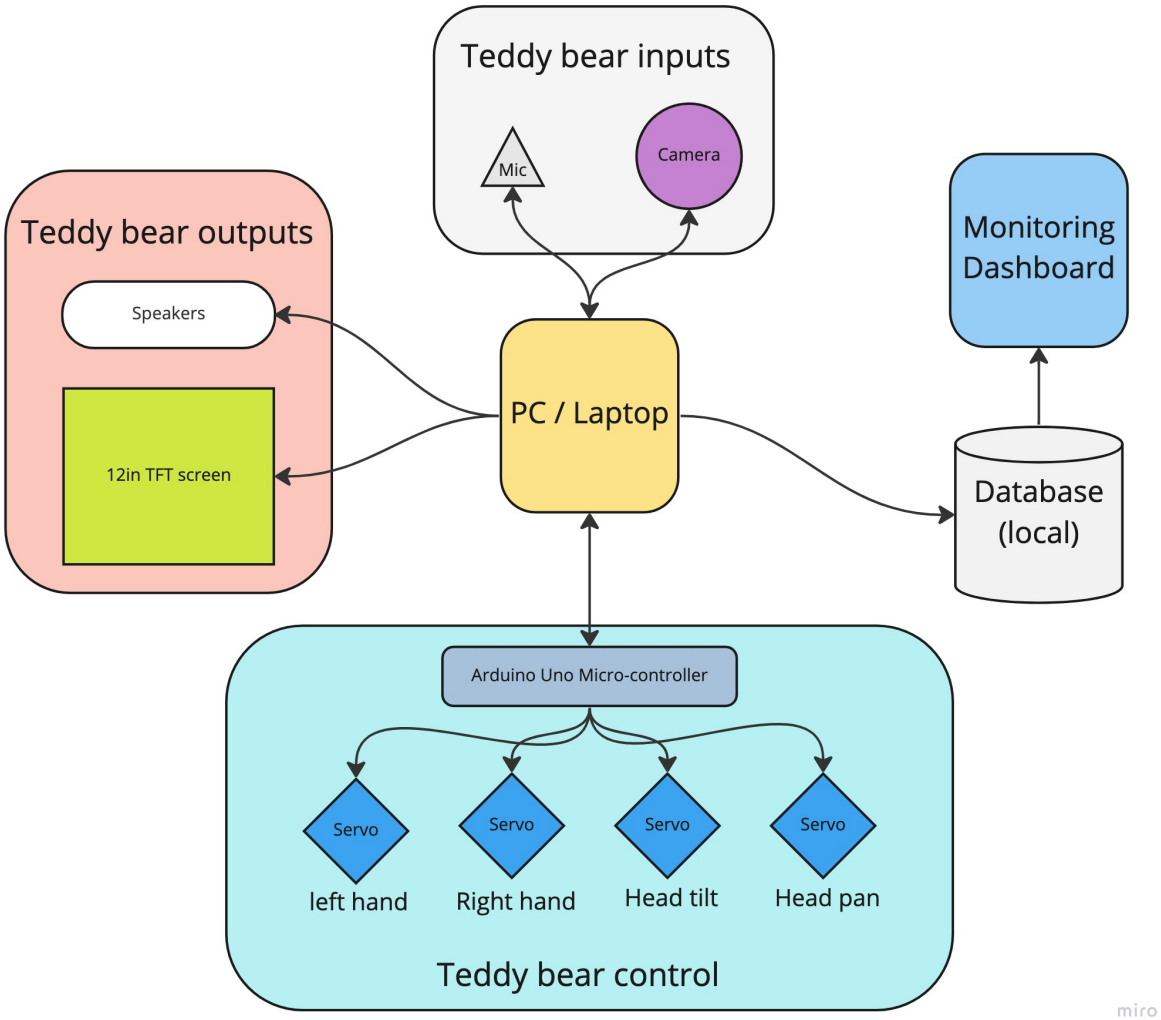


图2：硬件原理框图

Robofriend代码架构

为了运行和协调不同的算法、硬件及用户交互，我们开发了一套专有的 Python 软件栈，下面将对其进行简要介绍。代码可在 https://github.com/IdoMatan/ RoboFriend 获取。

代码架构基于 RabbitMQ。RabbitMQ 是一种类似 ROS 的发布者-订阅者框架，实现了异步并行进程控制。RabbitMQ 使我们能够同时控制机器人的伺服电机、摄像头、屏幕及其他所需外围设备，并运行我们稍后将介绍的算法。

图5展示了我们软件的示意结构，说明各个进程以及它们之间传递的消息。主进程是 StoryTeller，它负责协调其他进程之间的流程并显示视频。该进程运行一个循环：播放下一个故事片段，然后调用算法服务获取下一个动作。此循环会一直重复，直到故事结束。在整个循环过程中，机器人会转动头部，调整视角，使画面中的所有人脸都位于中央。图4展示了这一流程。

其他服务要么是基于定时器的（例如每隔 N 毫秒发送一帧），要么是事件驱动的，例如，页面结束消息会触发算法服务进行下一步动作计算。所有指标、动作和有用的元数据（不

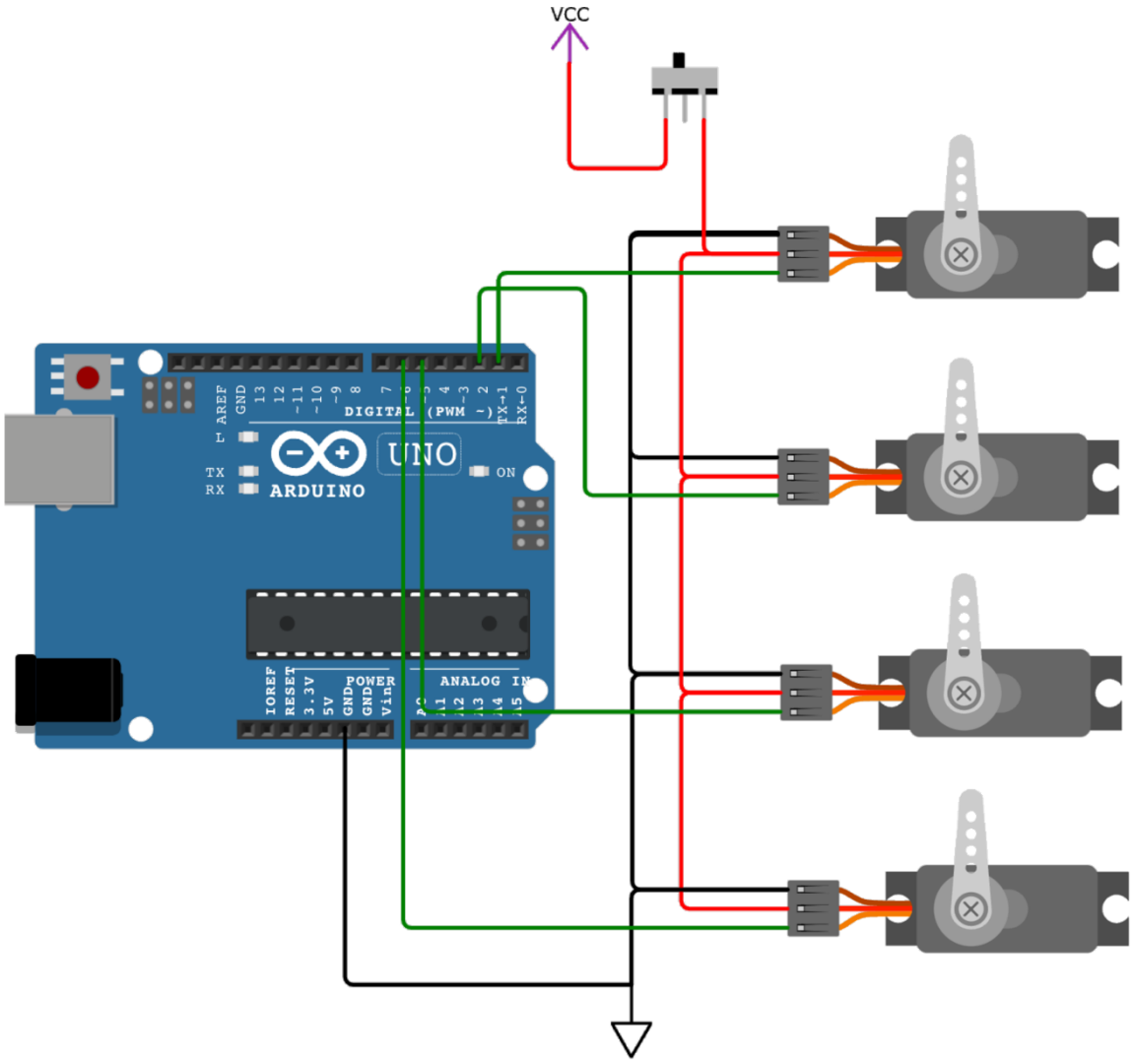


图3：Arduino 子系统

包括视频素材）都会实时记录到本地 PostgreSQL 数据库中，从而既可以在试验后进行分析，也可以通过 Grafana 仪表板进行实时监控。
图6展示了已实现的简易图形用户界面，用户可以在其中运行应用、选择一篇受支持的故事、选择一种操作模式（稍后将对此进行讨论），并开始或停止故事。

Robofriend算法

在介绍了 Robofriend 的机械结构和代码架构后，我们现在可以讨论用于控制它的算法。如前所述，我们的高层控制算法采用了带有“约束机制”的强化学习（Giacomo 等人 [2020], ），以避免机器人沿着我们已知不理想的轨迹行动。我们首先介绍感知和奖励函数，然后介绍机器人可执行的动作，最后说明在初步评估中用作“约束机制”的限制条件。

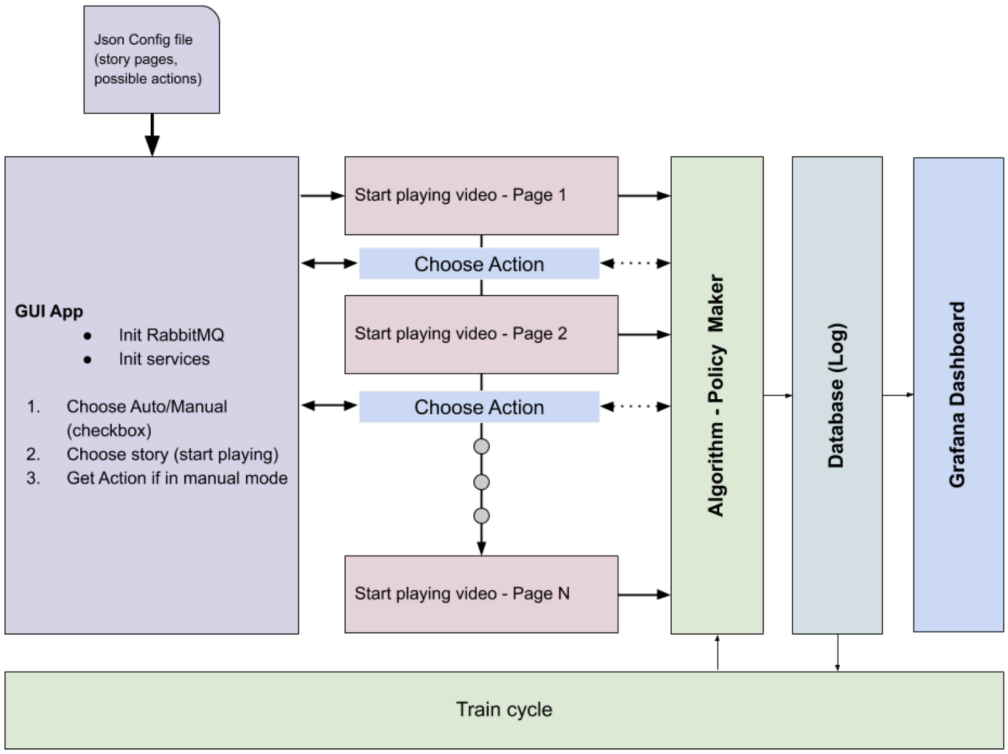


图4：软件运行流程

Robofriend 感知与奖励函数

如前所述，奖励基于计算机视觉，用于检测儿童的人脸和注视方向，并测量噪声水平。具体而言，我们使用 MTCNN 基于神经网络的人脸检测器来检测帧内儿童的人脸（Zhang 等人 [2016], ），随后使用 GazeNet（ Zemblys 等人 [2018] ）对每张人脸进行注视估计，从而生成相对于相机镜头的注视向量。由此得到以下指标:

人脸数量 MTCNN 人脸检测算法检测到的人脸数量。人脸数量发生变化，可能表示孩子走开了，或没有看向摄像头。

平均相对注视（注意力） 对于每张检测到的人脸（以索引 i 表示），GazeNet 算法会预测一个注视向量 θ_i, ϕ_i ，其中 θ_i 是水平（左/右）角度， ϕ_i 是垂直（上/下）角度——两个角度均相对于帧中心（相机镜头中心）。根据这些测量值，我们将奖励的注视 部分定义为:

$$r_{gaze} := \frac{1}{n}\sum_{i=1}^{n} cos(\theta_i)$$

粗略地说，我们的注意力指标范围为 0–1，反映儿童关注机器人的程度，而非环顾房间的程度。

兴奋度 通过连续帧，我们能够计算每张人脸的跳动程度指标，以反映儿童有多静止。这是衡量他们注意力与参与度的另一项代理指标。形式化地，设第一张图像中检测到的人脸（中心）位置为 x_i, y_i（对应 $i = 1 \dots n$），下一张图像中检测到的人脸位置为 x'_j, y'_j（对应 $j = 1 \dots m$）。由于人脸没有关联身份，我们必须先将第一张图像中的人脸与下一张图像中的人脸对齐。具体而言，对于第一张图像中的每个人脸位置 x_i, y_i，我们通过贪心方式，在下一张图像中寻找距离最近的人脸位置 x'_j, y'_j（使用欧几里得距离，且距离不超过可行的最大值）。然后，我们将人脸 i 的跳动程度定义为

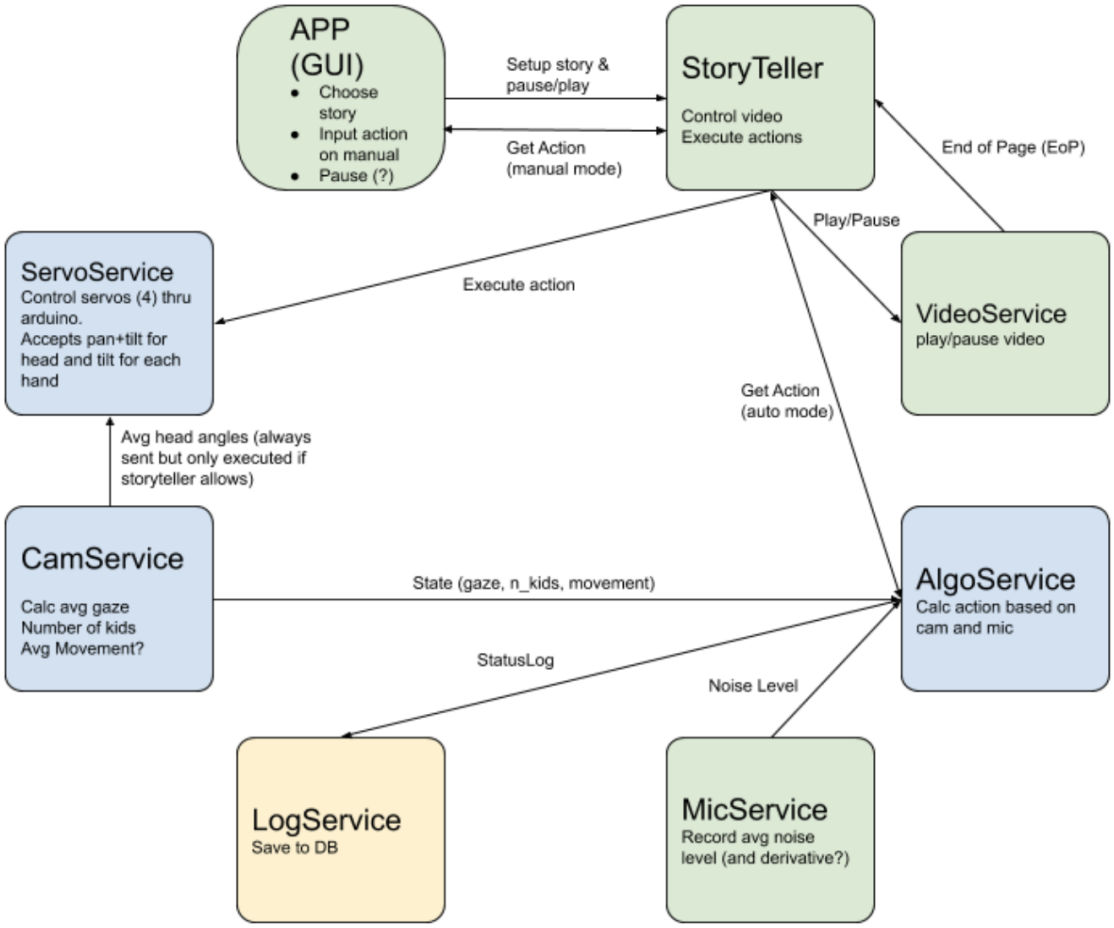


图5：软件架构图

该欧几里得距离，而总跳动程度则是每张人脸跳动程度之和，即 ：

$$r_{jump} := \sum_{i=1}^{n} \min_{j} \sqrt{(x_i - x'_j)^2 + (y_i - y'_j)^2}$$

噪声水平 由泰迪熊内部的麦克风测量，在 1 秒时间段内的平均噪声水平。这是儿童彼此交谈而不是聆听故事的程度的一项代理指标。我们用 r_{noise} 表示该指标。

噪声级别偏差 噪声水平的导数，同样按每秒计算平均值，用 r_{nd} 表示。设置这一指标的动机是捕捉一个页面内声级的变化，表明儿童可能由于前一项动作的影响而变得更加吵闹或更加安静。

为了汇总整个故事片段持续期间的奖励信号，我们对属于该片段的每个视频帧分别计算这些指标的平均值。请注意，我们使用平均值而不是总和，因为不同故事片段的持续时间不同。这样便得到状态向量 $\langle r_{gaze}, r_{jump}, r_{noise}, r_{nd} \rangle$。

最后，为了汇总这些 不同的指标并将其整合为单一奖励函数，我们采用加权求和，因此我们的奖励由 r 表示的 d 定义为

$$r := \alpha_1 * r_{gaze} - \alpha_2 * r_{jump} - \alpha_3 * r_{noise} + \alpha_4 * r_{nd} + ltl_reward$$

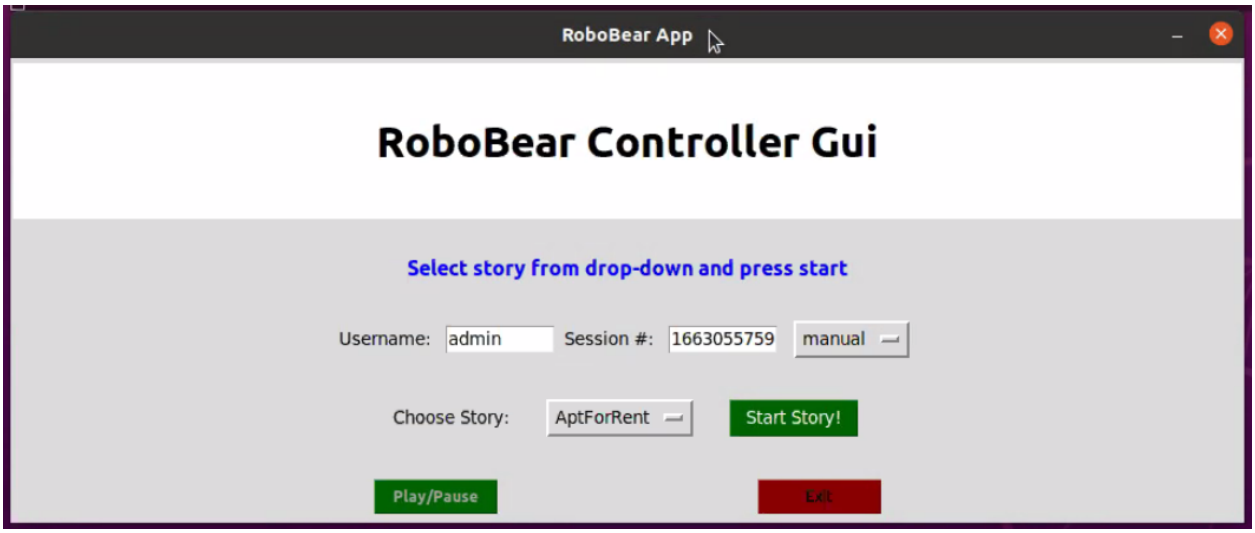


图6：用于与机器人交互的简单基于 Python 的图形用户界面

其中， ltl_reward 将在下文详细介绍；从概念上讲，它对应于一组机器人应学会遵守的预定义约束规则。

我们通过指出这些测量结果可作为真实奖励（儿童注意力）的代理指标来结束对奖励的讨论，而真实奖励无法直接测量。

Robofriend 的状态与动作

如上文所述，Robofriend 会阅读一个被划分为多个故事片段的故事，并在每个故事片段之后选择要执行的动作。需要注意的是，故事片段之间的顺序是线性的，因此无法选择接下来要阅读哪个故事片段。唯一的选择是每个故事片段之后要执行哪个动作。

Robofriend 的动作对应于它可以给予儿童的不同类型的反馈，分为：

积极反馈 此动作会从一组积极反馈语句中随机选择，例如“做得好”或“你听得真认真”。

消极反馈 此动作会从一组（轻微的）消极反馈语句中随机选择，例如“请注意听”或“请保持安静”。

提问 此动作会随机选择一个与 Robofriend 刚刚读完的故事片段相关的问题。Robofriend 不会尝试提取答案，而只是暂停适当的时间。

继续 立即继续阅读下一个故事片段。

移动头部和手臂 执行一系列随机的头部和手臂动作，持续几秒钟 .

请注意，所有这些动作无论 Robofriend 在故事中的哪个位置都适用（提问动作会随位置而变化，但始终适用）。此外，Robofriend 从传感器中提取的唯一信息就是奖励信号。因此，并不存在表示故事进展到何处的状态概念；唯一的状态信息就是上文所述的状态向量 $\langle r_{gaze}, r_{jump}, r_{noise}, r_{nd} \rangle$ 。不过，作为“约束机制”使用的线性时序逻辑约束确实包含自身的状态概念，我们将在下文讨论。

Robofriend 约束螺栓

Robofriend 控制架构的最后一部分是“约束机制”——机器人需要遵守的线性时序逻辑约束 Giacomo 等人 [2020]。也就是说，我们针对机器人允许和禁止的轨迹指定线性时序逻辑轨迹约束；在此情形下，轨迹是一系列动作。例如，我们可以指定如下约束，禁止连续提问：

$$G(ask_question \rightarrow X \neg ask_question)$$

（也就是说，如果当前动作为提问，那么下一步动作就不能是提问）。这可以确保机器人不会一直提问。

在对机器人进行初步评估时，我们测试了以下几种认为有必要纳入的约束机制，总体约束是以下各项的合取：

不要连续提问两次

$$G(ask_question \rightarrow X\neg ask_question)$$

不要连续挥手两次

$$G(wave_hands \rightarrow X\neg wave_hands)$$

最终要提出一个问题

$$F(ask_question)$$

最终要挥手

$$F(wave_hands)$$

Robofriend 学习算法

Robofriend 可以使用强化学习来寻找一种策略，在仍然有较高概率符合线性时序逻辑约束的同时，优化奖励的期望总和。具体而言，我们使用 Konda 和 Tsitsiklis [1999] 提出的演员-评论家强化学习算法，该算法在每一步输入儿童的状态，并预测下一步动作，以最大化整个回合中奖励的期望总和，同时考虑 Giacomo 等人 [2020] 提出的会修改奖励的线性时序逻辑约束。

鉴于此类算法通常需要大量且足够多样化的训练数据才能收敛，再加上具体问题设定的复杂性和训练资源的可用性，我们增加了一个选项，使系统能够根据外部“奥兹巫师”反馈，对策略网络进行模仿学习。该反馈由监督人员或教师实时提供。在机器人的首次试验期间，教师可以手动选择适当的动作，而算法会将其作为带标签的训练集，用于对策略网络进行初始训练。这不仅能为学习过程提供更好的起点，还能避免在训练过程最初阶段出现不希望发生的情况。

伦理考量

在构 建造 Robofriend 的过程中，我们遇到了许多伦理方面的考量。现在我们将对此进行详细讨论。il.

首先要考虑的是隐私问题，这可能会带来潜在风险，尤其是在涉及儿童时。因此，我们在设计 Robofriend 时，尽可能保障儿童的隐私。首先，Robofriend 不会保存任何录制图像。所有图像都会在线处理，以检测人脸（不会尝试将人脸与姓名关联起来——而且机器人本来也没有姓名）。Robofriend 记录的唯一数据，是每帧检测到的人脸数量，以及这些人脸的注视方向。

另一个潜在的顾虑是，Robofriend 会取代教师与儿童的人际接触。如前所述，Robofriend 是供托儿所教师使用的工具，让班级中的一部分儿童能够聆听故事，同时教师可以与其他儿童进行更多个别互动。这有助于教师为那些在有背景干扰时难以专注聆听故事的儿童投入更多个别时间，也能为那些在社会互动方面遇到困难的儿童补充相关活动。因此，Robofriend 是教学人员的辅助工具，而不是教师的替代品。

结论

我们介绍了 Robofriend 这款讲故事的机器人泰迪熊的构造和软件架构。未来，我们计划考察儿童对 Robofriend 的反应，以及它学习如何让儿童投入故事的能力。

参考文献

P. K. Kuhl. E arly language acquisition: cracking the speech code. 《自然·神经科学评论》, 5(11):831–843, 2004.

Lebdeva GC Cardillo and P. K. Kuhl. 婴儿语音感知的个体差异可预测儿童至 5 岁时的语言与阅读前技能。载于发展与行为儿科学会年会，2009 年。

C Moon、H. Lagercrantz 和 P. K. Kuhl。宫内经历的语言会影响出生后的元音感知：一项两国研究。《儿科学学报》，102(2):156–160，2013。

JS Hutton、T Horowitz-Kraus、AL Mendelsohn、T DeWitt 和 SK Holland。学龄前儿童的家庭阅读环境与听故事时的脑激活。《儿科学》，136(3):466–478，2015。E Twait、R Farah、N. Shamir 和 T. Horowitz-Kraus。学龄前儿童的对话式阅读干预与更强的认知控制有关：一项脑电图研究。《儿科学学报》，108(11):1993–2000，2019。Giuseppe De Giacomo、Luca Iocchi、Marco Favorito 和 Fabio Patrizi。强化学习智能体的约束机制。载于*AAAI*，第13659–13662页。AAAI出版社，2020。Kaipeng Zhang、Zhanpeng Zhang、Zhifeng Li 和 Yu Qiao。使用多任务级联卷积网络进行联合人脸检测与对齐。《*IEEE* 信号处理快报》，23(10):1499–1503，2016。Raimondas Zemblys、Diederick C Niehorster 和 Kenneth Holmqvist。GazeNet：使用深度神经网络进行端到端眼动事件检测。《行为研究方法》，2018。Vijay Konda 和 John Tsitsiklis。行动者-评论家算法。《神经信息处理系统进展》，12，1999。