《网络空间安全概论》课堂报告

汇报人：郭子铭

1. Agent行为指纹与轨迹关联

2. 犬类交流解码综述

Table 1: Related studies compared by target, dataset scope, observed evidence, and label or intervention. Their labels and candidate sets are not interchangeable.

Study	Target	Scope	Observed data	Label / intervention
System identity and workflow				
Trace [2]	Model family	7 families; 3 frameworks	Terminal commands	Model labels; passive traces and follow-on probe
ProcGrep [3]	Agent workflow	10 coding Agents	Coding actions and procedures	Workflow profile; passive traces
LIDAR [4]	Base model	36 models; 2 frameworks	Coding behavior	Model labels; active probes
Known By Their Actions [5]	Browser model	14 models; 4 environments	UI events and timing	Model labels; passive traces
AIT/AECID [1]	Recorded execution mode	Troubleshooting sessions	Commands, timing, syscall summaries	Human/Agent source labels; passive logs
Cyber-agent behavior and relation evidence				
CyBiasBench [14]	Attack preference	5 Agents; 3 targets	HTTP requests	Agent/target/attack-family conditions
Cross-Agent [15]	Campaign links	Synthetic sessions	Events, text, and timing	Synthetic campaign labels; logs and probes
Synthetic APTs [16]	Behavioral resemblance	1 model; 5 personas	Emulated TTP profiles	Controlled persona conditions
This study	Trace association	3 evaluation cohorts; 2 endpoints	Command sequences; observed launch gaps	Same-challenge proxy and source-mode labels



AttackAgent Detect



Behavioral Fingerprint Matrices for Execution-Mode-Aware Trace Association Across Security Tasks

AttackAgent Detect
Research manuscript in development

November 9, 2026

Round status: prior same-challenge retrieval results, an AIT mode-estimation pilot, and source audit are included.
The proposed K1–K8 matrix-specific association and independent-source transfer evaluations remain open.

Abstract

Behavior-based trace association uses tool choice and operation order to retrieve related executions, but recorded behavior can also reflect the task, model, harness, prompt, and collection interface. We present an observation-aware, fold-adaptive framework that selects available evidence and training-fold settings, then reports candidate ranking and link errors separately. The source-linked catalog contains 1,142 nonempty normalized records from Cyber Task Horizons and two Hugging Face releases, with source-qualified tasks and explicit observation contracts. In a separate same-challenge retrieval study across 17 CTH tasks, increasing the Agent share in constructed galleries shifted fixed-threshold errors from false merges toward missed links; matched CTH tests showed no learner-common BA gain from B/S. Frozen NYU Agent-only results were learner-dependent: XGBoost ranking improved, but its FMR reached 25.39%. An exploratory AIT pilot on 109 troubleshooting runs reached 64.4% balanced accuracy under source-group holdout with category-enriched B/S/T, compared with 85.4% for Command TF–IDF. Collection workflow remains a plausible source of label signal, so these results estimate recorded-mode labels rather than intrinsic actors. The proposed K1–K8 matrix-specific association and independent-source transfer evaluations remain open.

1 Introduction

Security investigations use behavioral evidence to retrieve relevant prior executions and relate observations across records. Some methods compare tool choices, local command combinations, and operation order. Agent participation introduces another influence on these features: independent activities can share an execution system, while one activity can switch between human-selected and Agent-selected operations. Similarity may consequently reflect shared models or harnesses, and differences may reflect a change of execution mode. Whether these effects increase false links or missed links is an empirical question, rather than an assumed monotonic consequence of adding Agents.

AttackAgent Detect develops a shared interface for behavioral observations, interpretable fingerprints, and association methods. Figure 1 summarizes the observation-aware pipeline and its separate mode-estimation and trace-association readouts. Association has several explicit endpoints: retrieving the same CTF challenge, matching a vulnerability target, or linking independently labeled activity episodes. These endpoints need their own labels. Task retrieval is a useful problem in its own right; it does not become operator attribution merely because the observations concern security tasks.

We propose a behavioral fingerprint matrix whose rows follow the execution and whose columns measure observable behavior. The representation preserves local changes in tool selection, action organization, and repetition. It can be used to estimate Human/Agent execution mode and can also enter an association method directly. Mode estimation need not be a mandatory intermediate decision or a rule that excludes cross-mode links.

A common representation is useful across diverse inputs, provided their observation contracts remain visible. Terminal-input captures, Agent shell-tool calls, and host execution logs expose different parts of behavior. Our corpus preserves those distinctions, task versions, and source provenance. Heterogeneous sampling therefore supplies multiple evaluation settings and failure conditions, rather than a single pooled identity benchmark. Evidence for a broadly usable tool must come from transfer evaluation on these settings.

We address three questions:

1

当攻击日志里混有人类与 Agent 的轨迹，行为画像能否分辨执行方式，
让同一模式下的同任务轨迹正确关联，同时把跨模式轨迹分开？

Table 2: Evaluation cohorts. Counts are eligible traces, not independent pair decisions; H and A denote source labels.

Source	Samples	Description
CTH mixed-role [11]	H × 34; A × 51	17 CTF tasks; five gallery shares for same-challenge retrieval. The normalized command views have no matched timing.
AIT audit logs [1]	H × 61; A × 48	109 runs across 12 troubleshooting tasks; recorded-mode recognition and launch-gap analysis. Durations are unavailable.
NYU D-CIPHER [13]	A × 585	585 Agent runs across 160 CTF challenges; frozen external same-challenge retrieval. Event timing is unavailable.

H 和 A 分别表示人工来源和 Agent 来源。CTH mixed–role 包含 34 条人工轨迹和 51 条 Agent 轨迹，覆盖 17 个 CTF 任务，用于同题检索；AIT 审计日志包含 109 次运行，来自 12 个故障排查任务，可分析记录模式和启动间隔；NYU D–CIPHER 包含 585 次 Agent 运行，覆盖 160 个 CTF 挑战，作为外部同题检索数据。需要注意，各数据集提供的时间信息并不相同，因此并非所有时序特征都能在每组数据上计算。

新形势下的关联分析挑战

- 内容：n-gram、embedding
- 序列：公共子序列
- 习惯：统计偏好
- 结构：子图匹配

噪声？特征！

谁在执行、用什么工具执行，从前在这里都属于该被抹掉的噪声

凡是经过，必留痕迹

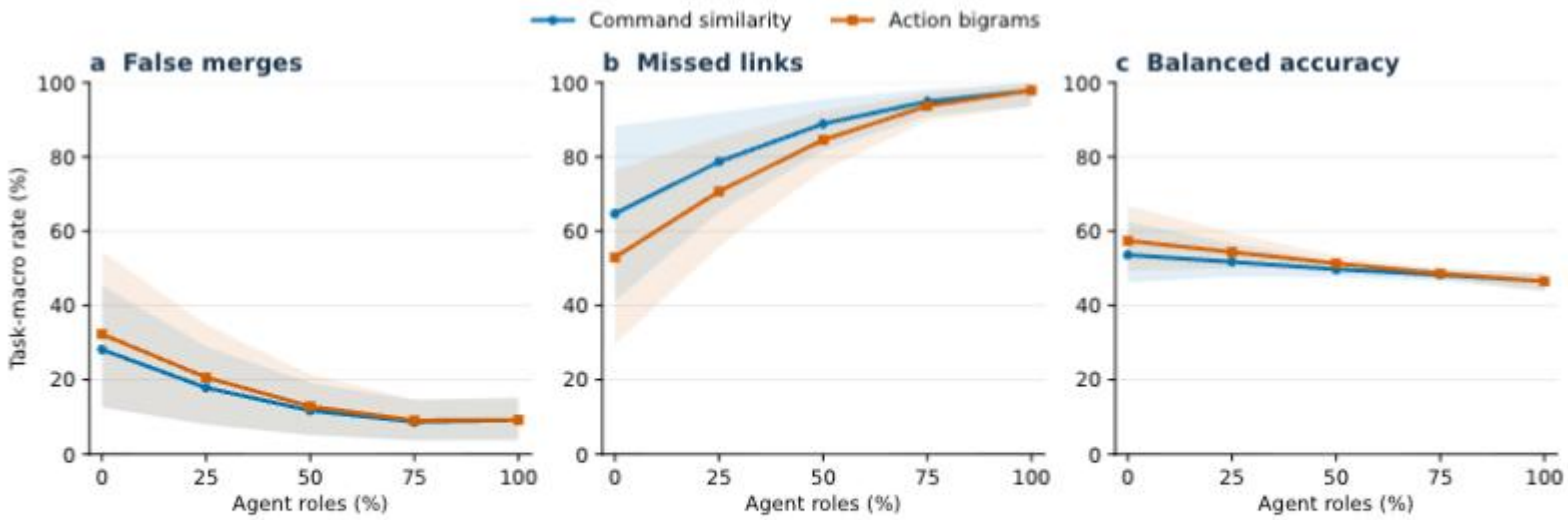
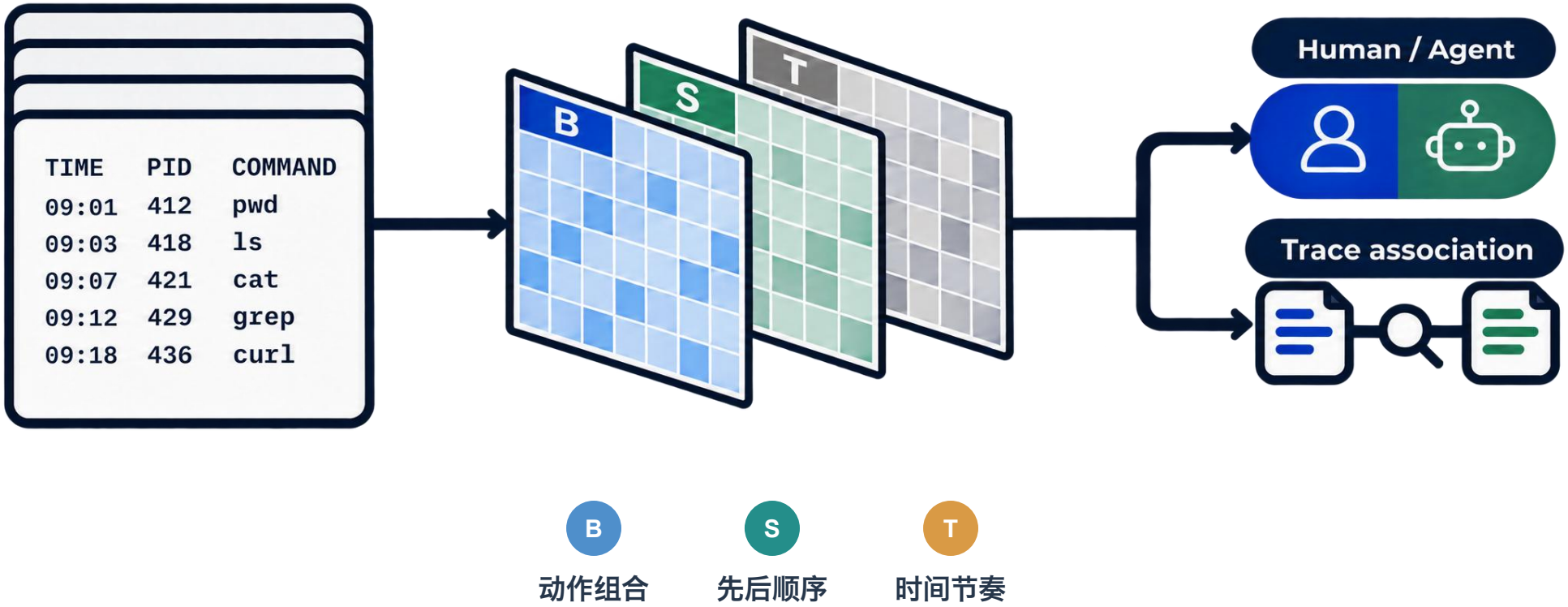


Figure 3: Command- and action-bigram-cosine baselines across five exact CTH gallery role shares. Panels show false merges, missed links, and balanced accuracy. Shading gives 95% task-cluster intervals; role share is a property of the constructed gallery, not a causal intervention on trace generation.

咋整？Agent的执行痕迹，也会留下AI的指纹

行为指纹由哪些线索组成?



例如：进程、文件与网络如何组合，关键动作怎样衔接，事件间隔和单步耗时如何变化。

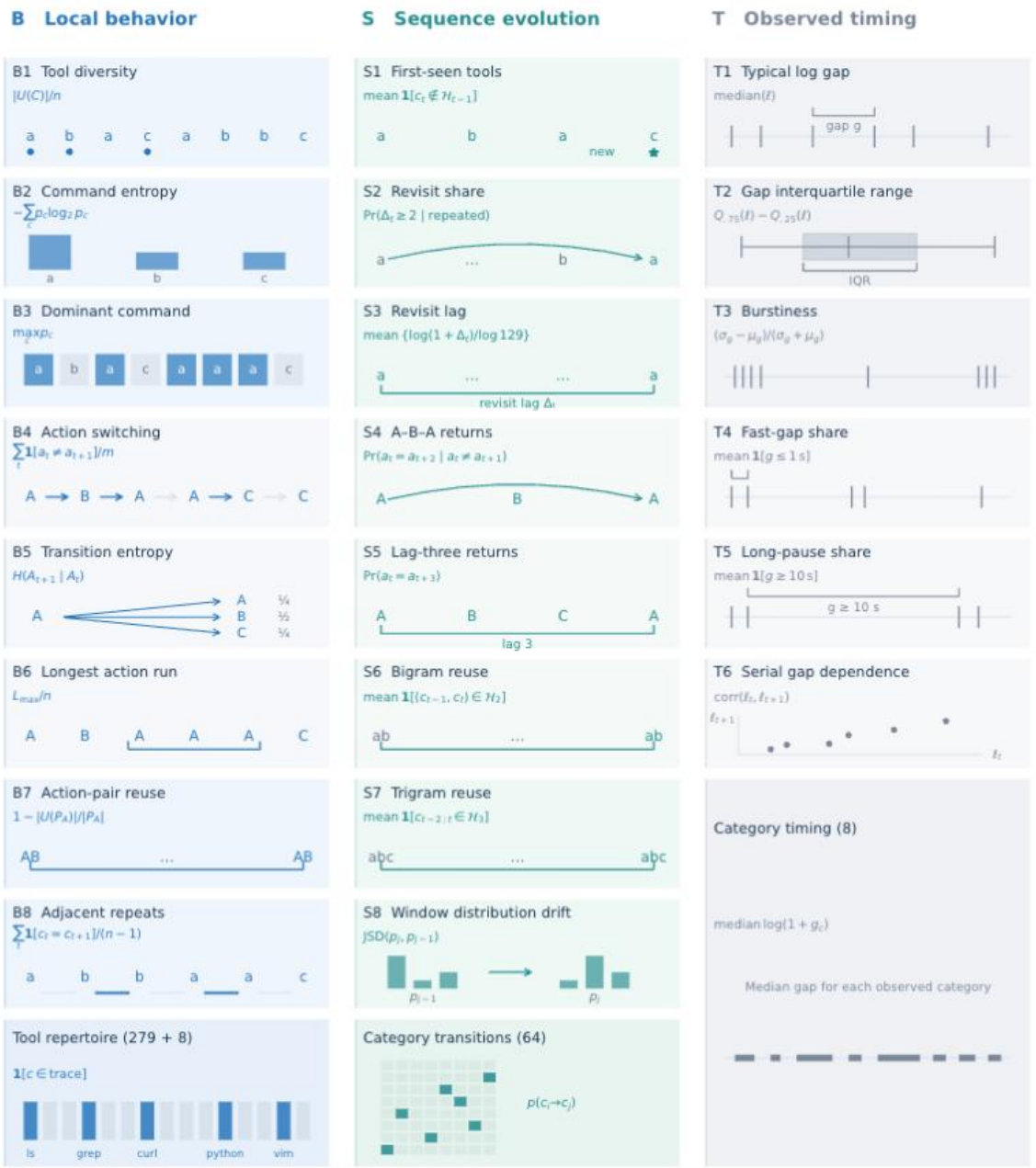
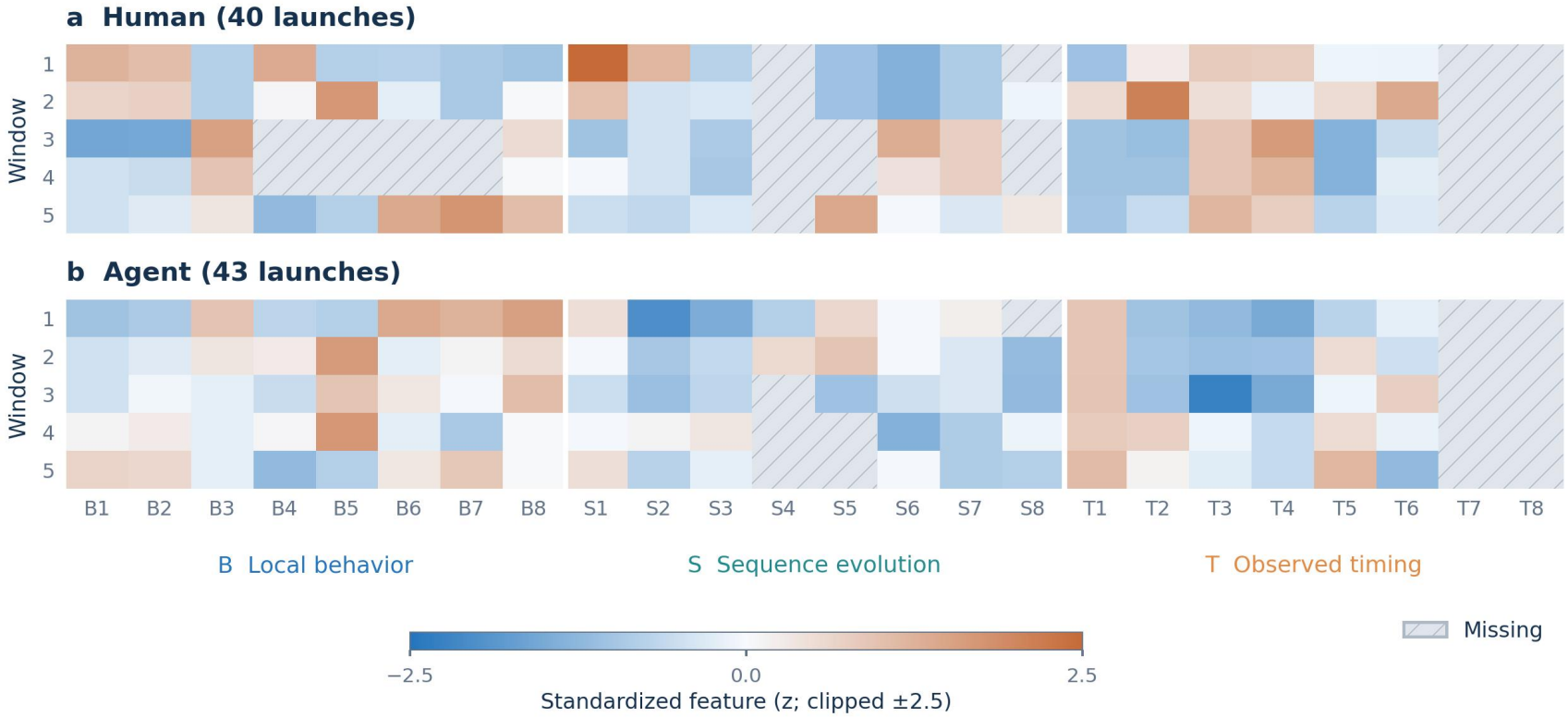


Figure 2: The 24-slot B/S/T candidate schema. Block colors distinguish local behavior (B), sequence organization (S), and observed timing (T). T1–T6 describe launch gaps; duration-dependent T7–T8 are optional and unavailable in AIT.

B：局部行为B1 Tool diversity使用了多少种不同工具；多样性越高，说明工具种类越丰富。
B2 Command entropy命令分布有多分散；使用频率越平均，熵通常越高。
B3 Dominant command最常用命令占全部命令的比例，反映行为是否集中在少数命令上。
B4 Action switching相邻动作发生切换的频率。
B5 Transition entropy执行某个动作后，下一个动作有多少种可能，反映动作转换是否多样。
B6 Longest action run连续重复同一动作的最长长度。
B7 Action-pair reuse同一对连续动作（例如 A→B）在轨迹中重复出现的程度。
B8 Adjacent repeats相邻两步动作相同的比例；图中用 a, b, b, a, a, c 示意。
S：序列演化S1 First-seen tools新工具第一次出现的情况，体现工具引入顺序。
S2 Revisit share对已用过的工具或动作进行回访的比例。
S3 Revisit lag工具或动作从首次出现到再次出现，间隔了多少步。
S4 A-B-A returns是否出现 A→B→A 这样的返回模式。
S5 Lag-three returns某个动作在间隔三步后再次出现的情况。
S6 Bigram reuse两步连续组合（例如 A→B）是否反复出现。
S7 Trigram reuse三步连续组合是否反复出现。
S8 Window distribution drift比较轨迹前后不同窗口中的动作分布，观察行为构成是否变化。
T：观测时序T1 Typical log gap相邻操作启动间隔的典型值；图中采用对数形式表示间隔。
T2 Gap interquartile range操作间隔的四分位距，反映间隔变化有多大。
T3 Burstiness间隔是较规律，还是集中成一阵一阵的“突发”模式。
T4 Fast-gap share间隔很短的操作占比，图中示意阈值为 1 秒。
T5 Long-pause share间隔很长的操作占比，图中示意阈值为 10 秒。
T6 Serial gap dependence前后相邻的两个操作间隔是否相关。

简单说：
B 看“做了什么、做得是否重复”，
S 看“按什么顺序发展、是否回访复用模式”，
T 看“操作之间隔了多久、节奏是否规律”。

同一任务的运行轨迹，也透露执行方式



B 动作组合

进程、文件、网络连接等行为
以及它们在同一时间窗里的共现

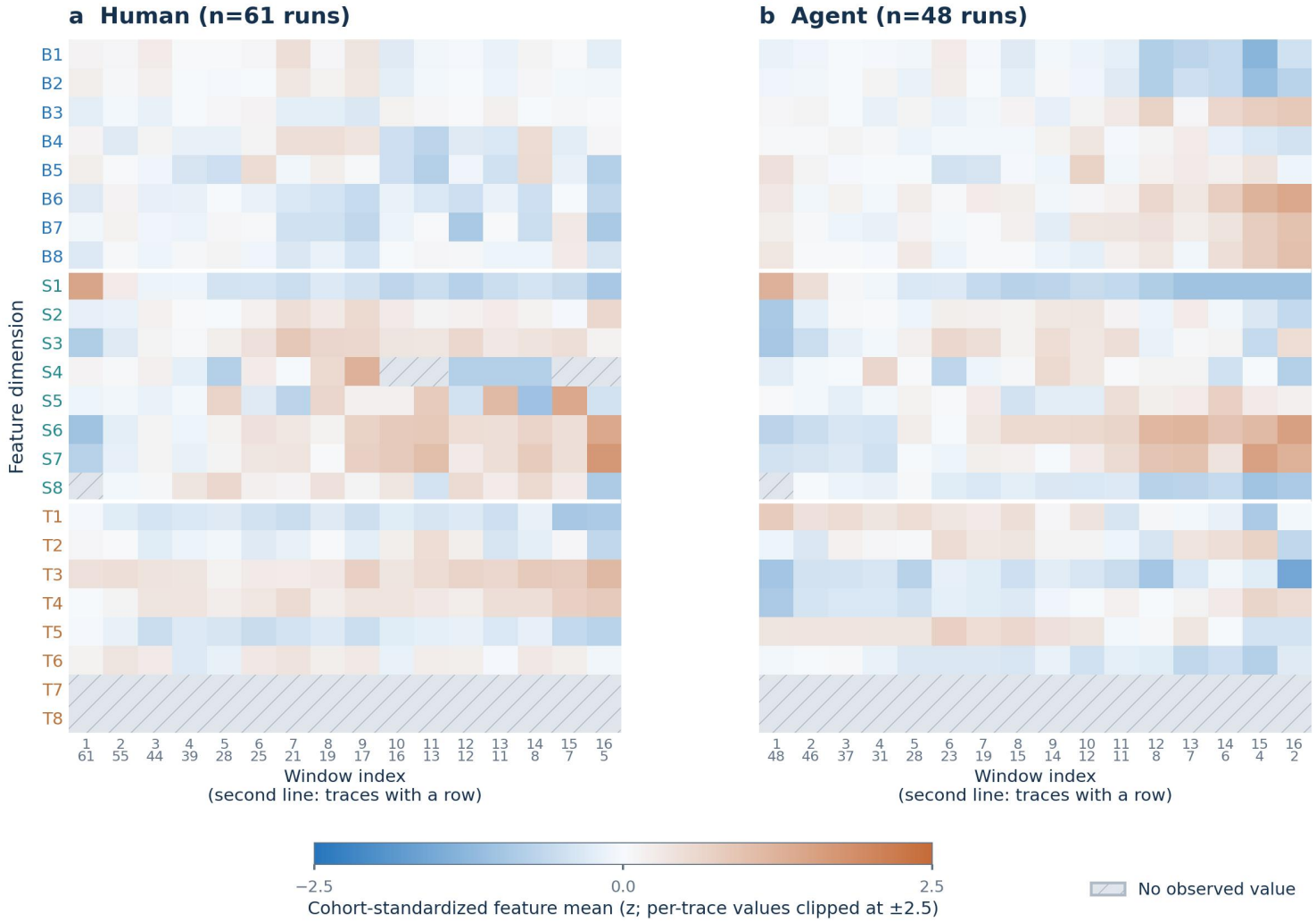
S 先后顺序

关键操作如何衔接、转向
以及路径分支怎样变化

T 时间节奏

事件间隔、单步耗时与操作频率
以及停顿和速度如何变化

主要实验结果1：AIT数据集样本均值观察



在前几个窗口中，两组较明显的差异体现在时序特征上：

人类组：
操作间隔更具突发性（T3），
且 1 秒以内的间隔占比更高（T4）；

智能体组：
中位间隔更长（T1），
10 秒及以上的间隔占比也更高（T5）。
S8 均值较低，说明相邻窗口间的动作类别分布变化较小。

主要实验结果2:

Table 4: AIT recorded-mode classification on 109 runs. Cohort counts are shown in Table 2; source holdout leaves entire actor groups unseen.

Method	Task holdout BA (%)	Source holdout BA (%)	Source AUROC
Log-length baseline	57.4 [44.2, 70.8]	32.9 [23.5, 43.7]	0.31
Legacy core-8	49.4 [38.5, 60.8]	39.3 [32.0, 46.0]	0.43
B+S	53.0 [41.0, 65.5]	35.8 [29.9, 42.4]	0.32
B+S+T	68.8 [61.8, 75.8]	52.6 [43.4, 62.7]	0.54
B+S+T+Category	89.4 [82.8, 95.3]	64.4 [55.6, 73.8]	0.72
Command TF–IDF	**95.1 [92.0, 97.9]**	**85.4 [77.4, 93.1]**	**0.94**

Balanced-accuracy intervals are 95% task-cluster intervals shown in square brackets. Teal shading identifies the fingerprint progression (light to mid) and the highest-scoring lexical reference (deep); all three columns are higher-is-better. Colors summarize point estimates and do not encode statistical significance. Source AUROC is the pooled source-holdout point estimate.

通过BST特征区分日志中 Human 和 Agent 执行模式的能力。完整指纹把轨迹整理成固定长度的结构化表示，包含行为、动作顺序、时间节奏、命令类别和工具使用。在任务留出测试中（换到新任务时），它的平衡准确率达到 92.6%，接近命令 TF-IDF 的 95.1%。

行为指纹既能把分散日志转成可比较的轨迹，帮助关联分析汇聚相关信息，也能提供区分 Human 与 Agent 执行模式的证据，让关联判断同时看见轨迹相似性和执行方式差异程度。

LIMITATION

我们的**“人 / Agent”二分简化了真实攻击场景，未能准确涵盖传统自动化脚本和人机协同；**
因为受限于现有数据，我们只能尽力解释可观测的轨迹差异，
不能据此严格判断实际执行主体。

Table 3: Combined feature-block ablation and candidate-ranking results for same-challenge retrieval.

Learner	Feature arm	Association decision			Candidate ranking		
		BA (%)	FMR (%)	FSR (%)	Hit@1 (%)	Gallery AUC (%)	Pair AP (%)
CTH mixed roles (17 tasks; 54,400 pair decisions per arm)							
RBF-SVM	**E (baseline)**	**52.57**	**6.52**	**88.35**	**33.75**	**58.73**	**39.78**
	E+B	54.08 (+1.51)	14.00 (+7.49)	77.83 (-10.51)	—	—	—
	E+S	52.45 (-0.12)	13.57 (+7.05)	81.54 (-6.81)	—	—	—
	E+B+S	53.95 (+1.38)	13.50 (+6.99)	78.60 (-9.74)	38.45 (+4.70)	62.22 (+3.49)	42.66 (+2.88)
Decision tree	**E (baseline)**	**52.80**	**6.59**	**87.81**	**34.91**	**60.30**	**35.21**
	E+B	50.31 (-2.49)	0.93 (-5.66)	98.44 (+10.63)	—	—	—
	E+S	51.79 (-1.01)	0.89 (-5.70)	95.53 (+7.72)	—	—	—
	E+B+S	50.00 (-2.80)	0.00 (-6.59)	100.00 (+12.19)	36.27 (+1.37)	59.77 (-0.53)	36.44 (+1.23)
XGBoost	**E (baseline)**	**56.06**	**5.97**	**81.91**	**34.40**	**58.92**	**40.97**
	E+B	52.75 (-3.31)	13.13 (+7.16)	81.37 (-0.54)	—	—	—
	E+S	56.10 (+0.04)	14.47 (+8.50)	73.34 (-8.57)	—	—	—
	E+B+S	53.68 (-2.38)	15.43 (+9.46)	77.21 (-4.70)	37.01 (+2.61)	61.73 (+2.81)	41.06 (+0.08)
AIT audit logs (12 tasks; XGBoost; 43,600 pair decisions per arm)							
XGBoost	**E (baseline)**	**50.12**	**0.00**	**99.76**	**40.20**	**63.34**	**38.25**
	E+B	50.94 (+0.82)	2.78 (+2.78)	95.34 (-4.42)	—	—	—
	E+S	53.18 (+3.06)	3.71 (+3.71)	89.94 (-9.82)	—	—	—
	E+T1–T6	52.42 (+2.30)	4.84 (+4.84)	90.32 (-9.44)	—	—	—
	E+B+S	50.94 (+0.82)	5.28 (+5.28)	92.84 (-6.92)	33.33 (-6.88)	57.87 (-5.47)	32.60 (-5.65)
	E+B+T1–T6	51.37 (+1.25)	5.24 (+5.24)	92.02 (-7.74)	—	—	—
	E+S+T1–T6	52.01 (+1.89)	3.53 (+3.53)	92.45 (-7.31)	—	—	—
	E+B+S+T1–T6†	51.92 (+1.80)	5.37 (+5.37)	90.79 (-8.97)	34.01 (+0.68)	58.80 (+0.93)	34.11 (+1.50)
NYU Agent-only external evaluation (frozen E+B+S versus E; 160 challenges)							
RBF-SVM	**E (baseline)**	—	—	—	**39.73**	**65.75**	**42.77**
	E+B+S	—	—	—	38.67 (-1.06)	65.56 (-0.18)	41.35 (-1.41)
Decision tree	**E (baseline)**	—	—	—	**36.92**	**63.94**	**35.29**
	E+B+S	—	—	—	31.33 (-5.59)	56.80 (-7.13)	32.29 (-3.00)
XGBoost	**E (baseline)**	—	—	—	**34.74**	**61.29**	**39.55**
	E+B+S	—	—	—	37.94 (+3.20)	63.10 (+1.82)	41.62 (+2.07)

Values are task-macro percentages; parentheses show paired percentage-point changes. CTH changes compare each arm with E. NYU ranking changes compare E+B+S with E. For AIT, non-dagger changes compare with E; the † row's ranking increment compares E+B+S+T1–T6 with E+B+S (paired decision increments are BA +0.98, FMR +0.09, and FSR −2.05 points). Higher BA and ranking scores are favorable; lower FMR and FSR are favorable. Teal/red fills show favorable/unfavorable point changes, with darker fills for larger changes; fills are descriptive and do not encode significance. Dashes mark metrics not included in that comparison. AIT measures launch-gap T1–T6 only. Complete intervals, denominators, thresholds, fallback counts, and fold settings are in the experiment record.

项目主页与使用示例

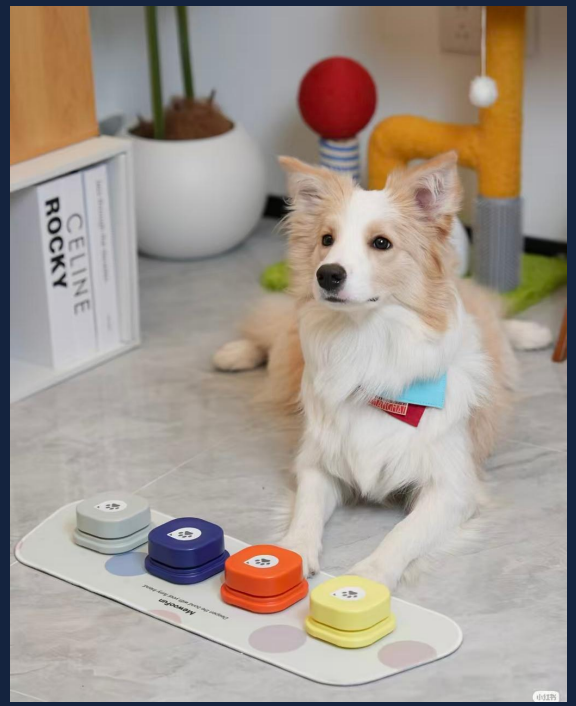


项目主页 · 示例

第二部分 · 犬类交流解码综述

一万四千年前，我们已经与狗并肩







约 1.42 万年前，一只幼犬与两位成年人共同安葬。

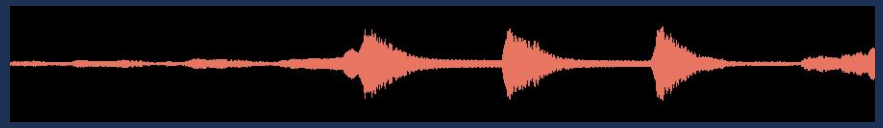
机器翻译的几次转向

1954	1966	1990	2014	2017	2020s
Georgetown - IBM	**ALPAC 报告**	**IBM Candide**	**Attention**	**Transformer**	**基础模型**
250 个词、6 条规则 现场俄英翻译演示	把质量、成本 和实际用途放到台面	从双语语料学习 统计翻译对应关系	让模型按上下文 学习输入输出对齐	纯注意力架构 提高并行训练能力	预训练表示迁移 扩展到更多任务

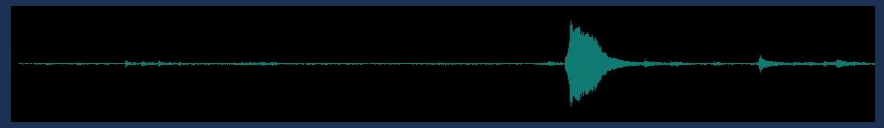
任务、数据、架构与评估，一起拓宽了机器翻译能做的事。

盲听：哪一段更像在玩？

A 点击试听



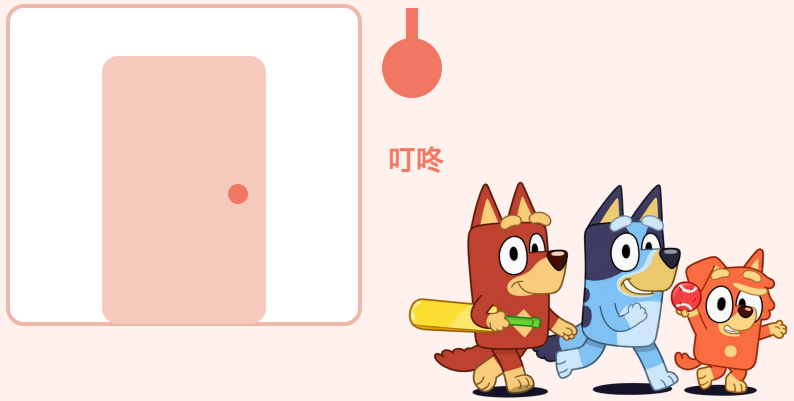
B 点击试听



举手投票：A / B

揭晓：A 门铃后，B 玩球时

A 门铃后的警觉



B 跟人玩球时的兴奋叫声



听出场景只是第一步：换只狗，还猜得准吗？

REPORTS

Word Learning in a Domestic Dog: Evidence for "Fast Mapping"

JULIANE KAMINSKI, JOSEP CALL, AND JULIA FISCHER Authors Info & Affiliations

SCIENCE • 11 Jun 2004 • Vol 304, Issue 5677 • pp. 1682-1683 • DOI: 10.1126/science.1097859

5,462 442 CHECK ACC

Abstract

During speech acquisition, children form quick and rough hypotheses about the meaning of a new word after only a single exposure—a process dubbed “fast mapping.” Here we provide evidence that a border collie, Rico, is able to fast map. Rico knew the labels of over 200 different items. He inferred the names of novel items by exclusion learning and correctly retrieved those items right away as well as 4 weeks after the initial exposure. Fast mapping thus appears to be mediated by general learning and memory mechanisms also found in other animals and not by a language acquisition device that is special to humans.

小狗理解我们在说什么吗?

瑞科（Rico，1994年12月13日– 2008年）是一只雄性边境牧羊犬，瑞科的饲主报告称其能理解超过200个简单的单词，于是莱比锡的马克斯·普朗克进化人类学研究所的动物心理学家茱莉安·卡明斯基（Juliane Kaminski）与同事对其进行了研究，这件事让瑞科上了新闻。卡明斯基等人在《Science》上写道，瑞科饲主的说法被证明是有道理的，40件物品中，瑞科平均能正确取回37件。瑞科在最后一次接触物品后，还能记住物品名称达四周。

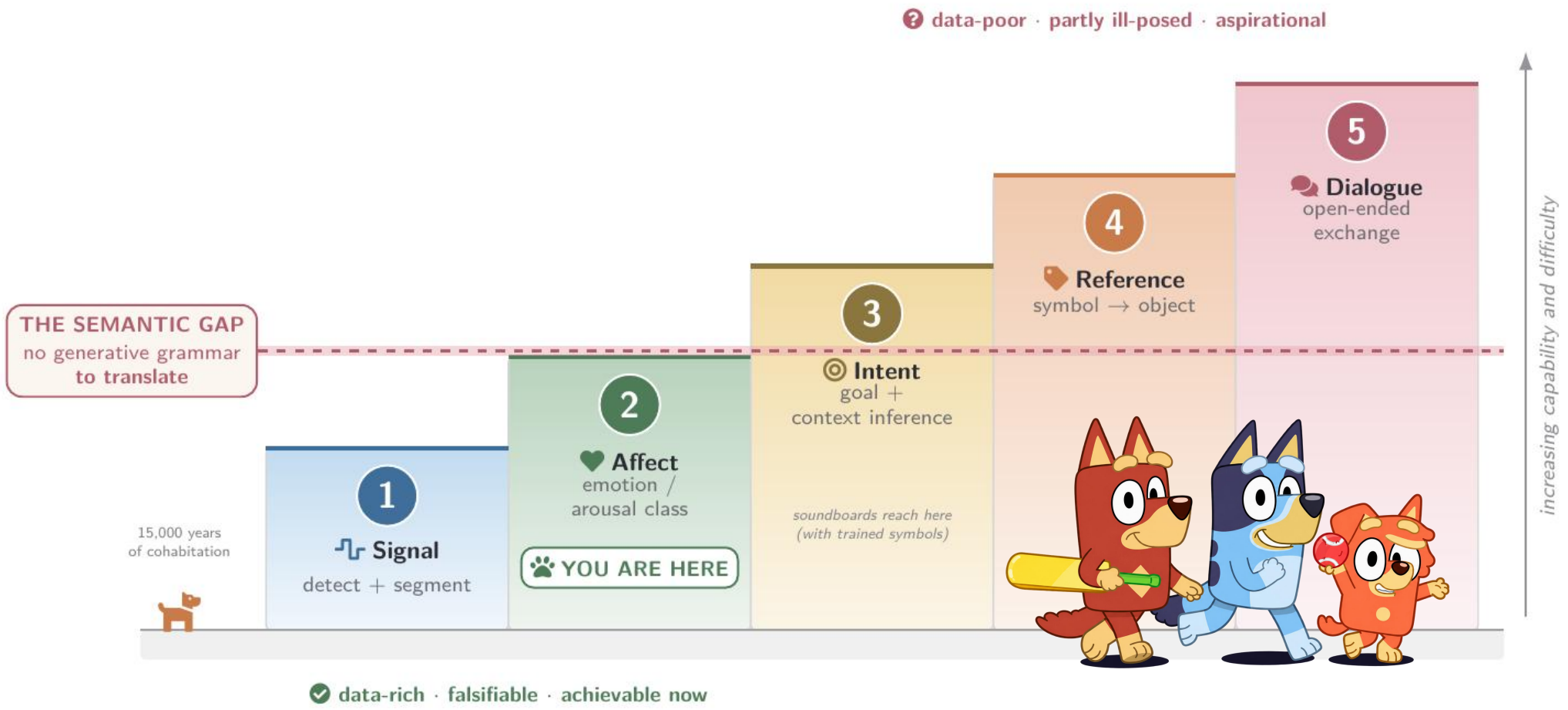
卡明斯基等人使用严格的试验计划排除了聪明汉斯效应：瑞科知道名字的200个物品，随机分配为10个物品一组，分配为20组。主人和狗先在隔开的房间里等待，实验者会在实验室内安置一组物品，然后再回到主人和狗的房间。接下来，实验者会指示主人要求狗从隔壁房间中（依次）带回两件随机拣选的物品，从而避免人类给出的其他信号的干扰。

犬吠研究：从借用语音表征到犬类专用模型

研究	方法	报告的进展	接下来检验什么
Abzaliev et al. LREC-COLING 2024	人类语音预训练 Wav2Vec2 微调犬吠数据	识别个体、品种、性别和场景；语音表示优于较简单基线	整只狗留作测试，检验新个体泛化
Dog2vec Interspeech 2025	HuBERT 式自监督学习 犬类声音专用预训练	6000 多小时视频清洗出约 150 小时犬吠；多个任务相对提升 8.2%	犬类专用表征能否覆盖不同品种与场景
NatureLM-audio ICLR 2025	音频语言基础模型 训练涵盖生物声学、语音与音乐	BEANS-Zero 多项任务达到 SOTA，包含未见物种零样本分类	迁移到犬类后，用行为记录核验场景预测

简单分类已经成熟，整句翻译还没开始

犬类交流解码的五级证据阶梯



犬吠模型还缺什么？

σ（sigma） ∈ S = Source，犬的信号集合 ℓ ∈ L = Language，人话句子集

encode → discretize → decode 声音 → 向量 → 单元 → 输出字符串	这步早有现成工具
D = {(σ , ℓ)} 要"这段叫声 = 这句话"的成对样本	世上也许不会有这种东西
ℓ 配对里的那一半：狗真正想说的话	没有任何独立途径看到它
score(f(σ), ℓ) 拿输出和它对照，判对错	没有对照物，分数无意义

出路只有一条，

第一，**降低对 ℓ 的要求**：

ℓ := 狗接下来做了什么（拿哪个物体 / 是否靠近 / 是否按按钮 / 按哪个按钮）

可观察 → score 恢复定义 → 但译出的是"意图"，不是"句子"

第二，**扩展 σ 的信号范围：（且看下文）**

机器翻译、脑机接口与生物声学的代表进展

领域	标志性进展	对犬吠研究的启发
机器翻译	1954 Georgetown - IBM 演示；1966 ALPAC；1990 IBM Candide；2014 Attention；2017 Transformer	任务定义、语料、架构与评估一起演进
脑机接口	1973 Vidal 提出直接脑机通信；2021 Willett 一位参与者手写脑机输入达 90 字符/分；2025 跨受试者迁移；2026 自定时静默语音控设备	报告跨个体迁移、校准成本和跨时稳定性
生物声学	2025 NatureLM-audio 探索跨物种音频迁移；2025 Dog2vec 训练犬吠专用自监督表示	比较通用音频迁移与犬类专用表示，再用行为证据核验

目前人类的脑机接口水平：
想一句话 → 读出词，12.5 万词表90.2%（使用8.4 个月后 97.5%，约 32 词/分钟），1 人校准成本半小时数据 → 50 词表 99.6%

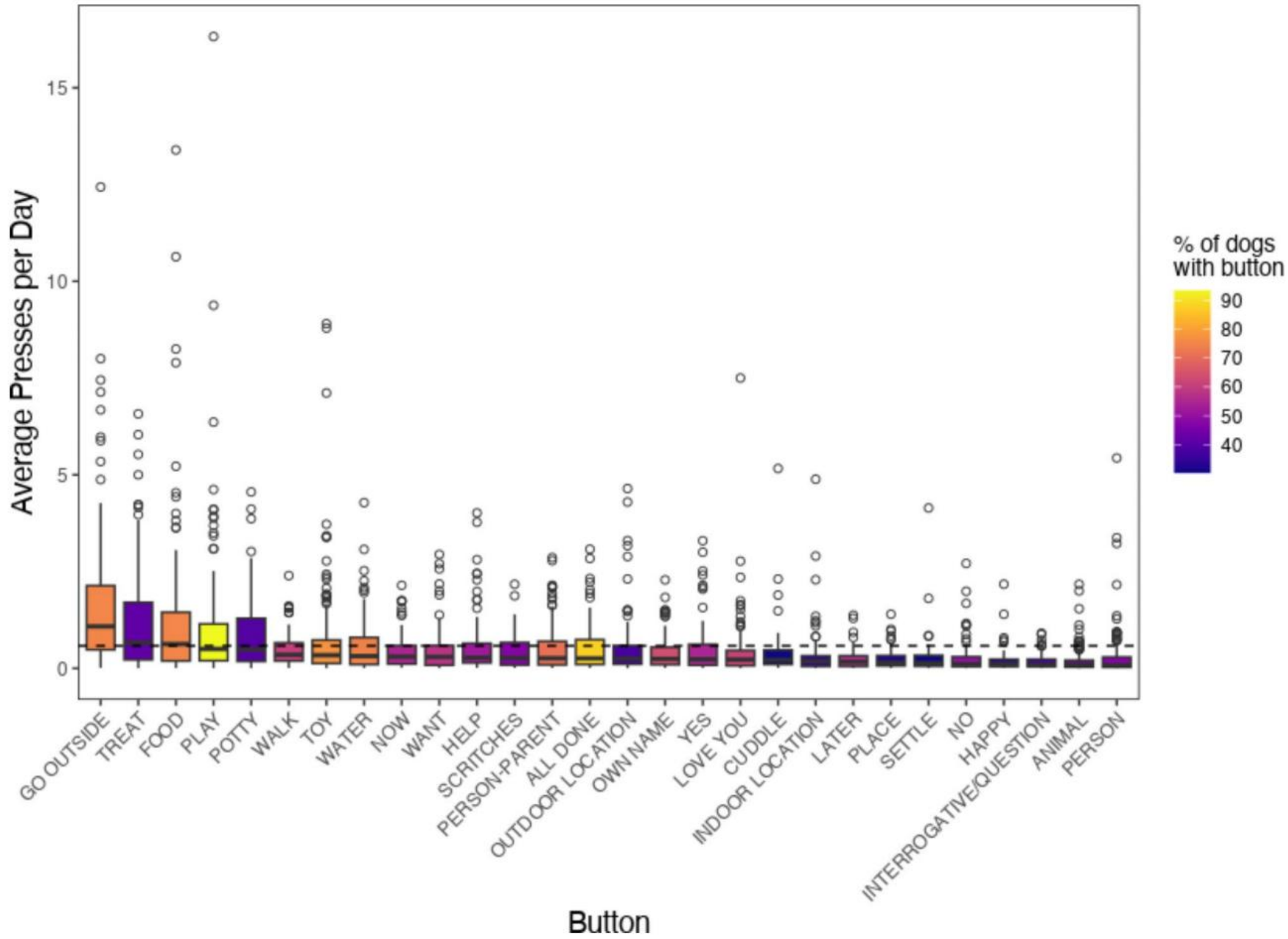


Fig. 1. Average daily dog pressing events for all button concepts for all 152 pet dogs. Colours indicate the percentage of dogs whose soundboards contained said buttons.

就像我们给llm工具，人类是时候也给小狗一个完整的输出接口了





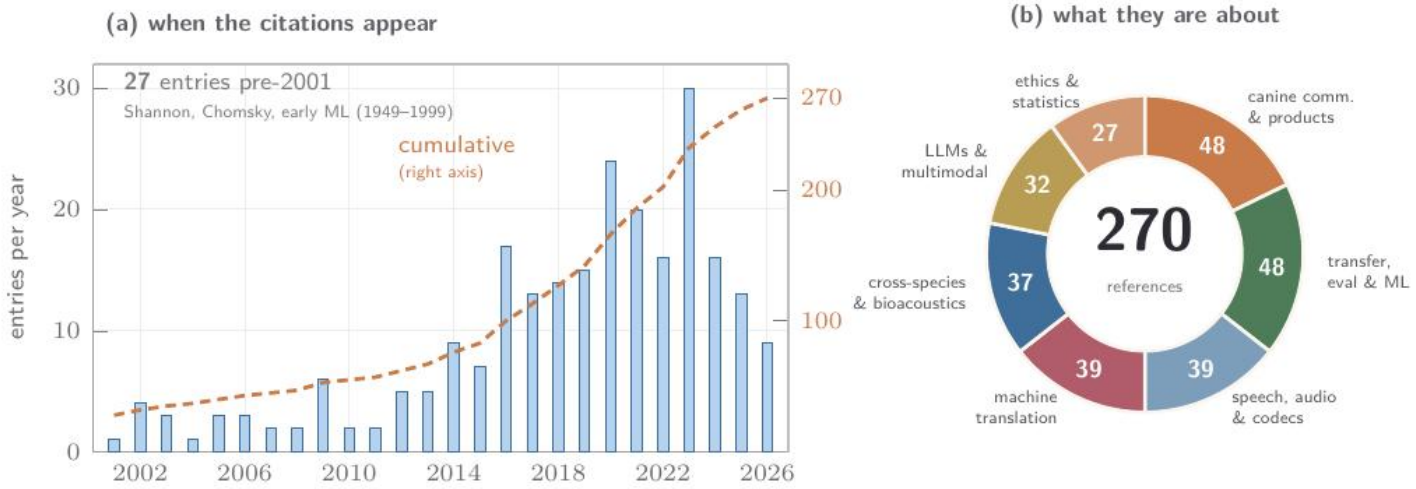
人非植入脑机范式能否迁移到小狗身上?

非植入范式	人侧怎么用	人侧准确率	能不能搬到狗
运动想象	指令："想象握拳 / 蹬脚 / 舌头 / 下颌"	四类公开基准 43.3%	**不能。**前提是先听懂一句关于内心的指令
P300 拼写器	6×6 格子轮流闪，你数自己想要的那个字母	健康人 90–99%，4–10 个正确字符/分	**能。**只需要注意，不需要语言指令
SSVEP / c-VEP	盯住以固定频率闪的图标，脑电里出现同频成分	分类 76–90%+，c-VEP 可达 140 bit/min（>20 字/分）	**最能。**非植入里速度最高的一档，只靠眼睛
fMRI 语义解码	听 16 小时故事训练"脑活动↔语义"映射，再让它想	重建意思，不逐字；250 选句式	**半能。**清醒犬 fMRI 已成熟，但要求几十分钟不动
Brain2Qwerty（*Nat. Neurosci.* 2026-06-29）	逐词看句子→记 1.5 秒→无反馈盲打键盘	MEG 平均字符错误率 29%（最好 18%）；EEG 65%	**不能。**它解码的是"敲键盘"这个动作本身

非植入脑机给狗用，可能实现，实现的是第三级——一条不用爪子的"选哪个"通道；但它也许永远实现不了第五级，因为第五级的东西不在信号里。



270 篇文献，拼出犬类交流解码的研究版图





The Canine Rosetta Stone woof-to-word

The Canine Rosetta Stone

How Far Are We from Translating What Dogs Say?

A Survey of Cross-Species Communication Decoding in the Age of Foundation Models

The woof-to-word Project

github.com/mcxiaoxiao/woof2word

November 2026 • Version 2.1

Abstract

Consumer apps already claim to translate dog barks into sentences, and flagship projects apply foundation models to whales, dolphins, and songbirds. How far are we, really, from translating what a dog says? This survey answers through a five-rung *translation ladder*: signal detection, affect classification, intent and context inference, referential semantics, and open-ended dialogue. Synthesizing canine behavioral biology, the human machine-translation stack, cross-species decoding projects, and the commercial market, we locate today's best systems on the second rung, with scattered footholds on the third. Two findings shape the verdict. First, a *semantic gap*: dogs signal affect, arousal, identity, and some referential content, but show no evidence of the generative syntax that sentence-level translation presupposes; above the lower rungs, "translation" is partly a category error. Second, an *evaluation crisis*: headline accuracies are unfalsifiable as translation metrics — no ground-truth intended message exists to score against. The lower rungs resemble heavier-than-air flight—hard but grounded, with existence proofs in hand—while open-ended translation of a non-linguistic signal resembles alchemy, whose target may not exist. The survey ships the instrument its diagnosis calls for: the Canine Ladder Benchmark, which fixes for every rung the error signal, the metric, and the baseline a system must beat, and a ten-item checklist applied to the claims the field has already made in public. We close with a capability roadmap, distance estimates, and an ethics framework that keeps the animal's welfare as the terminal criterion. Validated affect-and-intent inference is plausible within this decade: the achievable goal, and worth building.

CCS Concepts:

- • Computing methodologies → Artificial intelligence → Natural language processing
- • Computing methodologies → Machine learning
- • Human-centered computing → Human computer interface (HCI)
- • Applied computing → Life and medical sciences
- • General and reference → Cross-computing tools and techniques → Evaluation

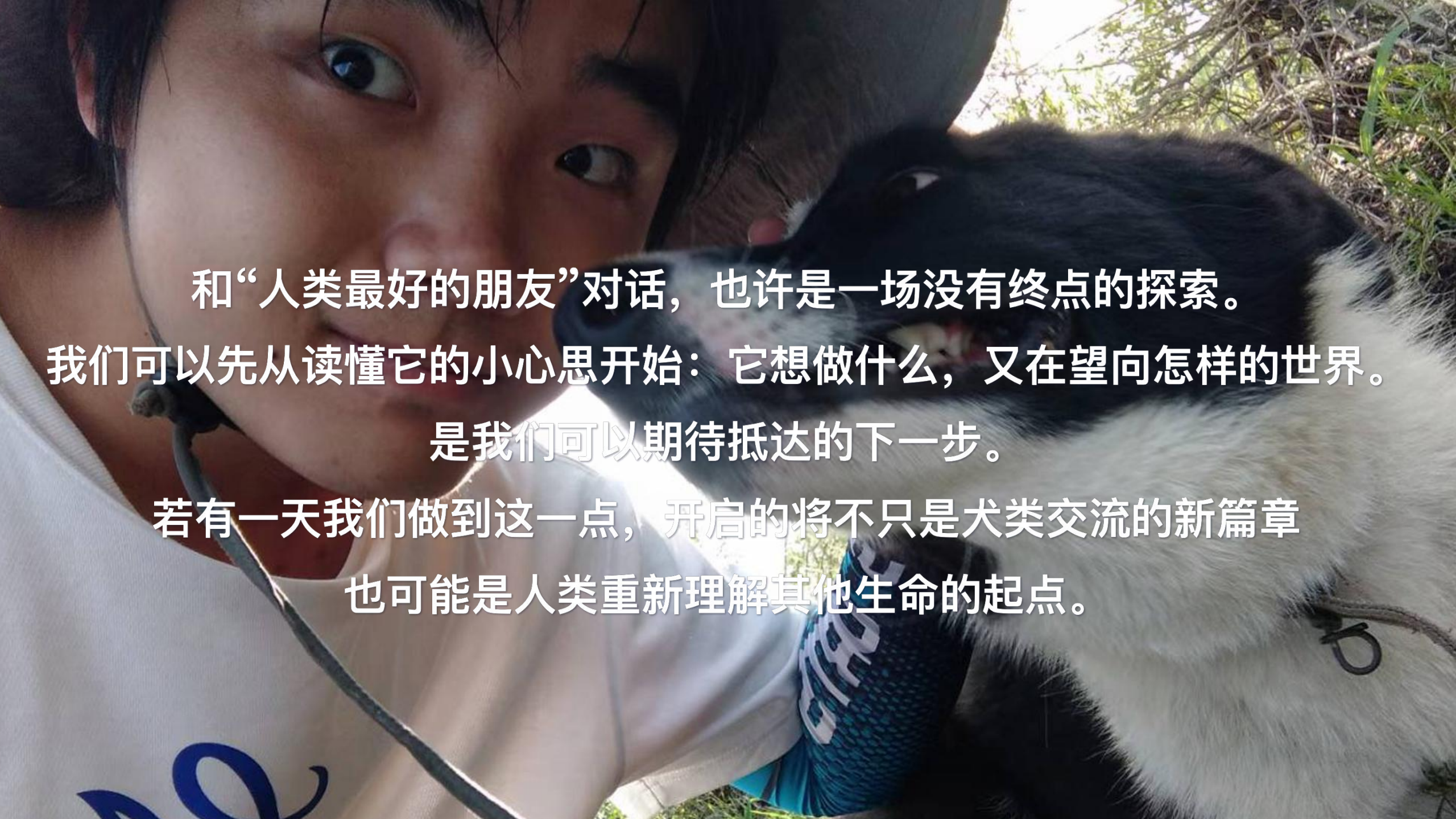
Keywords: animal communication, bioacoustics, foundation models, canine cognition, DogFACS, semantic grounding, interspecies interaction, evaluation, research ethics.

人类语音技术 → 跨物种迁移 → 犬吠分类

woof2word

github.com/mcxiaoxiao/woof2word





和“人类最好的朋友”对话，也许是一场没有终点的探索。

我们可以先从读懂它的小心思开始：它想做什么，又在望向怎样的世界。

是我们可以期待抵达的下一步。

若有一天我们做到这一点，开启的将不只是犬类交流的新篇章

也可能是人类重新理解其他生命的起点。