张志锐

zrustc11@gmail.com · (+86) 188-1063-6928 · zrustc.github.io

个人简介

深度机智 DeepCybo 合伙人、物理大模型负责人，统筹公司物理大模型方向的整体技术规划与研发。中国科学技术大学博士，MSRA 联合培养；具有大规模预训练、后训练与产品落地经验，发表论文 50 余篇，曾任 ACL 领域主席。
现聚焦物理智能与机器人智能，推动大模型走向真实世界的感知、推理与行动。

教育背景

中国科学技术大学 · 博士 2014 - 2019

计算机科学与技术，微软亚洲研究院联合培养。

中国科学技术大学 · 本科 2010 - 2014

计算机科学与技术。

主要经历

深度机智 DeepCybo 2026.05 - 至今

合伙人 · 物理大模型负责人

负责公司物理大模型方向的整体技术规划与研发，统筹模型、数据、训练和评测体系，带领团队推进具身理解、推理与机器人行动能力建设。

- **统一物理大模型**：带领团队研发 PhysBrain 1.5，将具身理解、动作生成和未来状态预测纳入统一自回归框架；8B 模型在技术报告的 28 项具身理解评测中均分 72.5，14 项居参评开源模型首位（含并列）。
- **人类示范学习与机器人落地**：推进 Human-as-Humanoid / PhysDex，将同步人类视频转化为 60 自由度机器人动作监督，实现叠杯、投环、倒水等任务在无目标任务机器人示范条件下的实机操作。
- **动作表示与评测体系**：推进 ActionPiece，探索保持物理邻域结构的动作 token 表示；建设 PhysBrainEvalKit，支撑空间与具身智能的可复现评测。

IDEA 数字经济研究院 2025.12 - 2026.05

基础软件实验室 · AI 技术顾问 / 创业合伙人

参与 MoonBit 的模型与工具全栈体系建设，协同代码数据、模型能力、工具链与开发工作流，探索更可靠的代码智能。

- 提出 SWE-AGI Benchmark，从编程语言、可验证执行和代码证明视角，探索代码生成与软件工程智能体的能力评估。

华为 2024.03 - 2025.09

终端小艺业务部 · 技术专家 A（20 级）

从零组建 10 人以上团队，负责小艺大模型通用后训练、推理模型及场景落地。

- **通用后训练体系**：牵引 SFT3.0 与数据、训练流程迭代，基于 Qwen2.5-7B 的方案在 AlpacaEval 2.0 和 Arena-Hard 提升 10%+，迁移至盘古 38B/135B-V3 后带来 10%+ 相对分位值提升。
- **推理与应用落地**：围绕快慢思考、可控思考长度及强化学习推进推理模型研发，支撑小艺深度解题智能体在 2025 电信打榜中获 AI 学习能力第一；推动对话 Live 态上线，RL + GenRM 将满意度从 3.35 提升至 4.15（+23%）。
- **多语言与多模态能力**：推动通话翻译中英双向平均提升 7 个 COMET 值；基于 Qwen3-8B 构建多模态后训练方案，通用能力持平或超过 Qwen2.5-32B-VL。

阶跃星辰 StepFun 2023.11 - 2024.03

算法专家 · 语言大模型预训练

负责语言大模型预训练及前沿技术探索，覆盖模型结构、训练效率与长上下文能力。

- 作为 Step2 近万亿参数 MoE 预训练项目核心成员，参与模型结构和训练策略验证；主导 FP8-LM 在预训练及后训练阶段的探索与落地，负责 LongContext 策略验证，支撑 128K 上下文能力。

腾讯 AI Lab 2021.09 - 2023.11

高级研究员（T11） · 自然语言算法平台组

负责腾讯翻译训练平台、交互翻译模型与个性化机器翻译，推动研究成果用于线上能力建设。

- **翻译平台与产品**：建设自动化数据及训练链路，搭建交互翻译多合一模型，支撑商业化交付；提出并落地个性化翻译策略 SK-MT，相关研究发表于 ICLR 2023。
- **多语言与多模态研究**：获 IWSLT 2023 端到端语音到语音翻译子任务第一名；参与混元大模型项目，负责多语言预训练数据清洗、翻译能力优化与 RLHF 调优。

阿里巴巴达摩院 2019.07 - 2021.08

阿里星 · 算法专家 / 语言技术实验室

负责阿里云通用翻译模型、语音翻译与商业化项目，贯通模型研究、训练平台与线上服务。

- 建设自动化训练链路，完成通用模型优化与语向扩展（15 至 214），线上解码加速 2-3 倍；开展多语言翻译、预训练融合与领域自适应研究，提升线上及定制模型效果。

微软研究院 2015.07 - 2019.06

研究实习生 · MSRA / Redmond Deep Learning Group

在 MSRA 及 Redmond Deep Learning Group 开展机器翻译、对话系统与自然语言生成研究。

- 提出 Iterative Back-Translation、Target-bidirectional Agreement 等方法；参与 Babel 项目，推动中英新闻翻译达到人类可比水平。

代表性论文

1. **PhysBrain 1.5: From Vision-Language Models to Physical Foundation Models**. *arXiv 2026*.
2. **EvoSpatial: Experience-Driven Skill and Tool Evolution for Spatial Reasoning Agents**. *arXiv 2026*.
3. **ActionPiece: Rethinking Action Tokenization for Autoregressive Vision-Language-Action Models**. *arXiv 2026*.
4. **Human-as-Humanoid: Enabling Zero-Shot Humanoid Learning from Ego-Exo Human Videos with Human-Aligned Embodiments**. *arXiv 2026*.
5. **SWE-AGI: Benchmarking Specification-Driven Software Construction with MoonBit in the Era of Autonomous Agents**. *arXiv 2026*.
6. **Is PRM Necessary? Problem-Solving RL Implicitly Induces PRM Capability in LLMs**. *NeurIPS 2025*.
7. **Safe Delta: Consistently Preserving Safety when Fine-Tuning LLMs on Diverse Datasets**. *ICML 2025*.
8. **Document-Level Machine Translation with Large Language Models**. *EMNLP 2023*.
9. **Simple and Scalable Nearest Neighbor Machine Translation**. *ICLR 2023*.
10. **Achieving Human Parity on Automatic Chinese to English News Translation**. *Technical Report 2018*.

完整论文列表：zrustc.github.io/publications.html

技能、学术服务及获奖

- Python / C++ / CUDA / PyTorch / TensorFlow；大规模训练、后训练与多语言模型系统。
- ACL 领域主席；ACL、EMNLP、NAACL、AAAI、IJCAI、NeurIPS、ICML、ICLR 等会议长期审稿人。
- 国家奖学金（2013）、谷歌奖学金（2013）、MSRA 实习生“明日之星”（2019）；腾讯 2 次 SEVP 卓越个人。