

# 腾讯 AI 大模型专家学习笔记

本纪要包含一部分： Q&A

本纪要所有数据及信息以上市公司公告口径为准

## 摘要：

- 1、 腾讯 AI 大模型六个 BG 都有参与, 其中 TEG 在算法上更偏 general, 其他 BG 更偏行业。
- 2、 混元大模型万亿参数水平, 相比 ChatGPT 性能更强, 但可能比 GPT4 比有所不足。混元大模型目前偏自然语言, 而 GPT4 支持多模态。
- 3、 推理过程边缘侧和训练侧是两套系统, 是分离的。
- 4、 目前来看通过翻译扩充中文语料提升不会很大。
- 5、 大模型之后会接入微信、游戏、短视频、广告、TOB 端等业务。
- 6、 腾讯目前部门向上申请的资源中浪潮占比偏多。
- 7、 目前数据均采用联邦学习计算方式进行加密, 不会导致原始数据泄露。
- 8、 精调不需要更新全部网络, 不需要更新后面几层, 故算力和数据要求较小。

## 第一部分：Q&A

### Q：腾讯在 AI 大模型方向的进展情况？混元大模型？

A：腾讯的大模型情况与百度、阿里不同，我们分到了每个 BG，我们有 6 个 BG（TEG 技术中台、CSIG 面向 B 端、微信、游戏、PCG 做内容、CDG 偏广告金融），其中六个 BG 都有做 AI，都非常重视不管从投入还是老板重视程度来看，我认为 AI 在技术侧是腾讯 number one 的地位。。TEG 类似中台，AI lab 在 TEG 下面。混元大模型由腾讯唯一 17 级科学家张正友老师负责，也是 AI lab 的负责人。6 个 BG 都有人在参与混元，共同推进，算法上 TEG 偏 general，其他 BG 偏行业。大模型腾讯在几年前也做过，混元也有上一代产品，偏检索

式。马上五月要发的是偏深层次的，对标 ChatGPT。目前混元还在测试研发中。

**Q：腾讯 AI 大模型在什么样的水平？**

A：腾讯 AI 大模型是万亿（参数）的，虽然还未发布，但我们自己认为比 ChatGPT 要强，但比 GPT4 可能不足，GPT 是支持多模态，而我们暂时偏自然语言。

**Q：混元大模型的算法是腾讯原创？还是基于什么技术路线？**

A：算法上对标 GPT，但也有不一样的地方，比如 Fine-tune、Reward Model、PPO，但其实这三步都是大家需要做的事情，主要是需要自己找一些新的数据，再做数据预处理。

**Q：在算力层面，腾讯云已用于 AI 训练的算力在什么水平？后续投入计划？**

A：我们投入了将近万张卡，我们在算力上还是有一定储备。我们也在内部评测华为昇腾芯片，发现有些场景下性能还可以。

**Q：混元的训练数据以什么类型为主，数据来源？**

A：数据来源为公开中文数据集、自己内部数据。数据需要做标注等预处理工作。

**Q：能否介绍一下您正在研发的产品具体情况？未来商业化场景？**

A：我们主要面向 B 端、G 端客户，后续会重点推行业的中小模型，主要有几种模式：1）对于超大客户，可能会选择从头训练一个大模型，但这种情况比较少，我们有能力去帮助客户从 0 到 1 重新训练大模型。2）基于客户的私有数据，对我们基础大模型做定制化 fine-tune。3）大多数客户还是会基于其行业属性，我们一起联合训练模型。

**Q：腾讯混元大模型在哪几个方面的性能强于 ChatGPT？**

A：从参数上，我们 1 万亿，ChatGPT 是 1750 亿。

**Q：腾讯的 AI 服务器用哪家较多？**

A：浪潮。

**Q：腾讯在 AI 医疗方面有没有与企业合作做 AI 医药研发训练，或者 AI 在线诊断？AI 医疗数据有专门数据处理公司吗？**

A：我跟的领域没有做这方面的事。我知道有一个单独的团队在做 Drug AI 的事情，但实际做到什么程度我不清楚。也有团队在做医保卡相关的项目。数据处理公司也不太清楚。

**Q：整个训练过程都是有监督条件下的大模型训练吗？**

A：前面模型可以做自监督，后面的要在模型上做数据标注。算法原理是先有一个一般意义上的模型，再在模型上做数据标注、排序等。

**Q：在混元大模型上做行业小模型，小模型属于推理部分吗？**

A：大多数客户只要推理，但费用和投入肯定不一样。

**Q：在推理过程中，数据传输和算力只有大模型需要还是推理端也需要？**

A：取决于客户部署在哪里。如果部署在公司云上，需要采购公司算力。如果有自己的私有云，就有自己的算力。针对数据隐私公司会加密。

**Q：如果应用端是 b 端企业，跑行业数据会落在边缘侧和训练侧不联系吗？**

A：是完全隔离的，可以认为是两套系统。训练后会出参数，之后进行部署和训练是分离的。

像在现在体验的 ChatGPT 只会用推理部分，与训练无关。

**Q：以混元大模型为基础开发的行业小模型在算法上要与大模型适配？**

A：网络结构肯定要一致。本质上都是神经网络，训练输出权重，推理就是利用网络结构做矩阵层。

**Q：结构一样可以迁移吗？**

A：每家都是完全不同的，像安卓安装的应用要重新开发才能安装到 ios 系统。

**Q：英文、中文语料在 AI 大模型上的差异？有可能通过翻译扩充中文语料吗？**

A：现在更缺的不是量，而是高质量数据。翻译本身会有信息损耗，目前英文比中文多了几倍，目前翻译不见得有很大的提升。

**Q：知识性语料翻译损耗低？文学类相对高？**

A：知识性英文语料在中文百科已经有了，没必要翻译。搜索频率极低的也没必要翻译。西方文学名著应该都有翻译了，非常小众的可能没有。

**Q：中国处于数据安全不太可能完全接纳 ChatGPT？微软在 bin 和 office 上的 Chatgpt 能否通过技术优化跨过监管要求？**

A：除非后续 OpenAI 同意把数据都放在政府监管之下，可能才能被容纳，这与技术没有关系。

Bin 和 office 应用同样，微软要接受政府监管，技术上像 VPN 是不会统一的，因为从个人安全、国家安全来看，数据是重要资产。

**Q：训练大模型时间？大概多久超越 GPT4？**

A：之前有过底座模型，现在是基于底座模型研发，不是从零开始。时间将近 5 月份，近小半年时间。

**Q：百度、华为、阿里、小米等技术层面对比？**

A：大家方法肯定类似，只是每家数据有差异。像百度做搜索引擎的肯定有先发优势，后续还要看把技术优势商业化能力。

**Q：腾讯模型后续进入方向？**

A：1) 微信肯定是最好的场景，因为它本身偏对话具有小程序，后续加上服务肯定有更好的体验。2) 游戏上肯定设计偏创意设计，效率上肯定有提升。3) 广告业务也会有很好的工具赋能。4) 短视频肯定也会有很大提升。5) TOB 上肯定会往各行业推大模型，因为 AI 本身是降本增效的工具。

**Q：浪潮服务器采购量？**

A：服务器采购量不知道。但申请资源里浪潮偏多。

**Q：大模型商业化会率先 To B 还是 To C？**

A：同时进行，因为每个 BG 均有参与，同时基于其底座能力。

**Q：目前是否有对外合作？**

A：目前 SaaS 加速器、微信等业务均有大量合作伙伴。原业务基于 AI 的部分合作业务均可以使用大模型进行替换。

**Q：大模型合作中的数据保护形式？**

A：客户行业数据主要用于训练部分对模型进行定制。目前数据均采用联邦学习计算方式进行加密，不会导致原始数据泄露。联邦学习原理可以实现只共享数据权重，数据和权重之间是单项传递，无法倒推。

联邦学习分横向联邦、纵向联邦。联邦学习根据不同用户的权重，将训练出来的模型加权平均，再分布给用户进行下一轮训练。

**Q：云端精调收费方式？**

A：采纳腾讯 IaaS 算力进行训练，腾讯也有自己的特殊数据帮助客户训练。

**Q：精调与预训练两个过程对芯片要求是否有区别？**

A：预训练主要是通过反向传播对全量网络进行更新。但精调不需要更新全部网络，不需要更新后面几层，故算力和数据要求较小。

**Q：有一定算力基础的厂商是否可以自己进行精调？**

A：腾讯万亿量级底座效果会较好，但目前斯坦福也已经推出了小参数但效果不错的模型。

伴随算法不断迭代，未来算力不会是瓶颈，数据才更为重要。

**Q：若使用昇腾芯片，会通过使用华为云还是采购芯片第三方代工方式？**

A：从技术来说是都可行的，但是从成本上来说，华为云成本会比采购更低。但腾讯肯定不会使用华为云，会使用自建云。对客户来说，还需要考虑的方面是数据安全性。但将数据交给国内大厂都是较为安全的。

**重要声明：**

《证券期货投资者适当性管理办法》于 2017 年 7 月 1 日起正式实施。通过微信、文档等形式制作的资料仅面向专业投资者，请勿对本资料进行任何形式的转发。若您非专业投资者，为保证服务质量、控制投资风险，请勿使用本资料中的任何信息。因本资料难以设置访问权限，若给您造成不便，烦请谅解！感谢您给予的理解和配合。

本资料所有者不会因接收人收到文档而视其为客户。对本资料内的信息的准确性和完整性不作任何保证，也不保证所包含信息和建议不发生任何变更。已力求文档内容的客观、公正，但文中的观点、结论和建议仅供参考，不包含作者对证券价格涨跌或市场走势的确定性判断。文档中的信息或意见并不构成所述证券或相关证券的买卖出价或征价，投资者据此做出的任何投资决策与本公司和作者无关。

本资料所载的资料、意见及推测仅反映本公司于文档当日的判断，本资料中所涉及到的证券或投资标的的价格、价值及投资收入可升可跌，过往表现不应作为日后的表现依据；在不同时期，本公司可以发出其他与本资料所载信息不一致及有不同结论的报告；本资料所反映的不同观点、见解及分析方法，并不代表本公司或其他附属机构的立场；本公司不保证本报告所含信息保持在最新状态。同时，本公司对本资料所含信息可在不发出通知的情形下做出修改，投资者应当自行关注相应的更新或修改。

本公司及作者在自身所知范围内，与本资料中所评价或推荐的证券不存在法律法规要求披露或采取限制、静默措施的利益冲突。

本资料版权仅为本公司所有，未经书面许可，任何机构和个人不得以任何形式翻版、复制和发布，且不得对本资料进行有悖原意的引用、删节和修改。未经授权刊载或者转发本资料的，本公司将保留向其追究法律责任的权利。