

信创筑基 智算未来

# 2025金融业新一代数据中心发展论坛

2025.09.12 | 贵阳

主办单位：



金科创新社

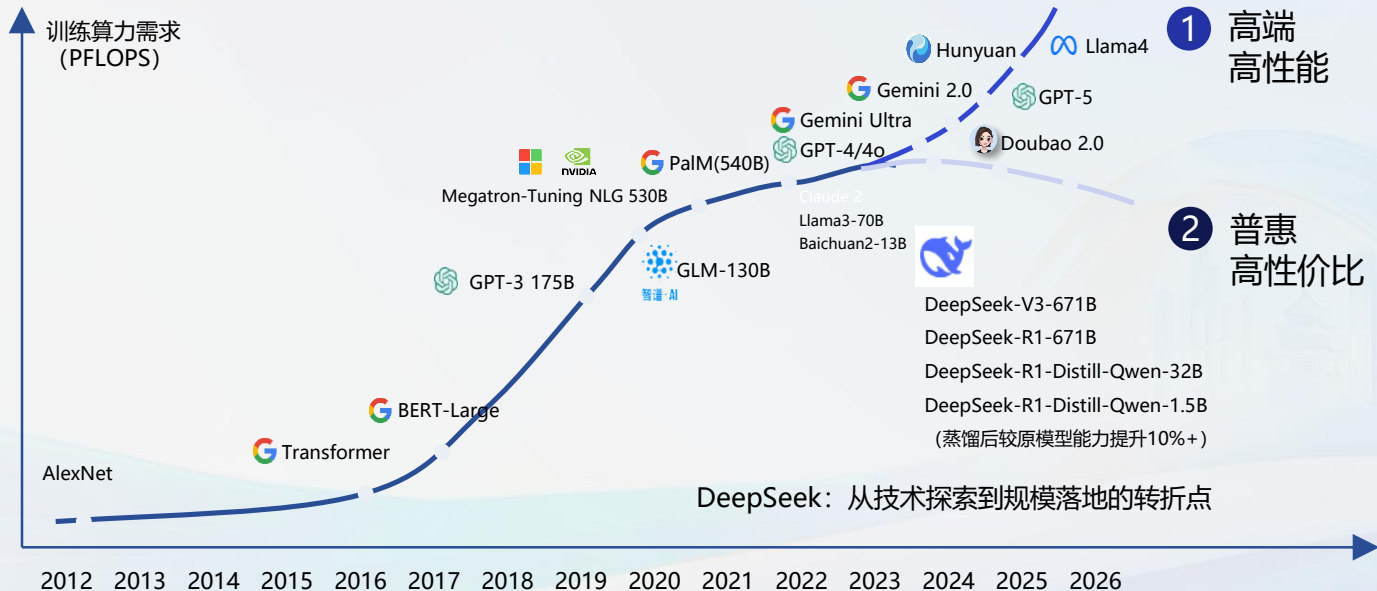
Fintech Innovation in China



# AI智算引擎助力金融行业效能跃升

王永超 | 联想金融行业资深架构师

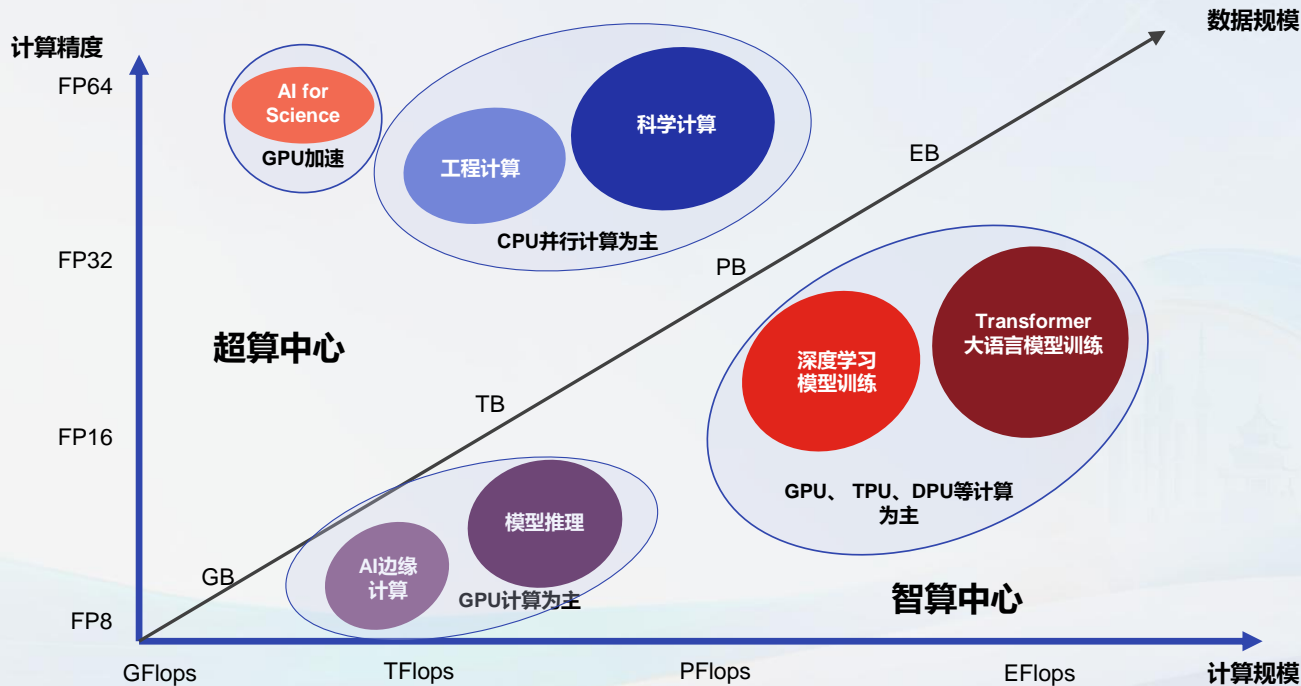
# 大模型2.0的新赛道



1 高端  
高性能

2 普惠  
高性价比

# 超算与智算的差异



# AI在金融行业的应用场景

智能体覆盖金融业务全链路，前台精准服务、中台实时风控、后台自动合规，实现端到端智能化重构。



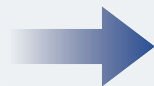
◆ 营销智脑：客户服务及流失预警

◆ 风险智脑：风险预测及投研决策

◆ 运营智脑：智能合规及技术升级

## AI技术创新不止，用户需求演变不息

有限的基础大模型开发用户  
(100+)



广泛的千行百业AI用户  
(100,000+)

**预训练**  
高MFU 高可用

综合  
全面

- **训练** 高MFU
- **训推** 高可用
- **推理** 高吞吐 低延迟

# 联想智能体矩阵赋能金融业务场景

前台智能服务

## 客服智能体（魔方）解决方案

“魔方”提供全渠道接入、AI辅助与情绪识别，实现客服效率与体验双提升

省时

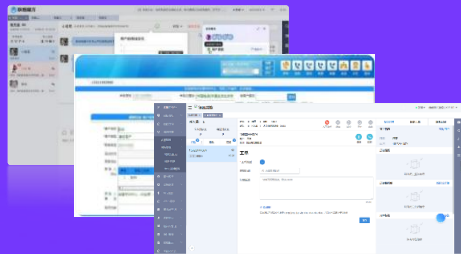
端到端客服方案

省力

一体化AI升级

省心

数据安全有保障



90%+

客户满意度提升

180%+

坐席效率提升

75%+

运营成本降低

中台智能决策

## 智能投研解决方案

通过智研、智问、智绘三大能力，构建企业级投研中台，全面提升投研效率

智研

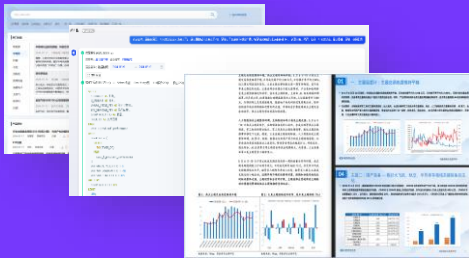
智能体解读研报

智问

智能问数工具

智绘

智能报告生成



100%+

投研覆盖率提升

200%+

数据准确度提升

90%+

创作效率提升

后台智能合规

## 智能行为合规解决方案

运用多模态AI技术实时监控销售合规，灵活部署助力风险防控

实时

多模态合规监测

精准

一体化智能风控

全面

全流程合规保障



90%+

合规检查效率提升

95%+

风险拦截准确率提高

80%+

人工审查成本下降

# 大模型应用面临的挑战



# 大模型训推部署架构



# 异构智算平台在AI计算中的作用

AI软件生态：工具、模型 ...



AI算力生态：芯片、核心部件 ...

# 联想万全异构智算平台持续升级

联想以前沿技术为用户实现定制化服务优势

## 大模型预训练



### 算力匹配魔方

为用户跳过繁杂的算力选择和验证



### 超智融合调度

提升训练效率  
节省额外支出



### AI高效断点续训

提升训练效率  
节省额外支出



### 联想集合通信库

AI训练效率  
提升10%-15%



### GPU内核态虚拟化

虚拟化GPU算力利用率  
从80%提升至95%

差异化技术进一步升级

## 大模型训练 + 训推 + 推理

### AI推理加速算法集

AI推理性能提升5-10倍

### AI编译优化器

训练和推理计算开销各降低15%+

### AI训推慢节点故障预测与自愈系统

百卡秒级、千卡分钟级、万卡十分钟级自愈

### 专家并行通信算法

推理延迟降低3倍+

# AI推理加速算法集 - 推理性能提升5-10倍

相比业界最优社区方案保持**20%+**优势

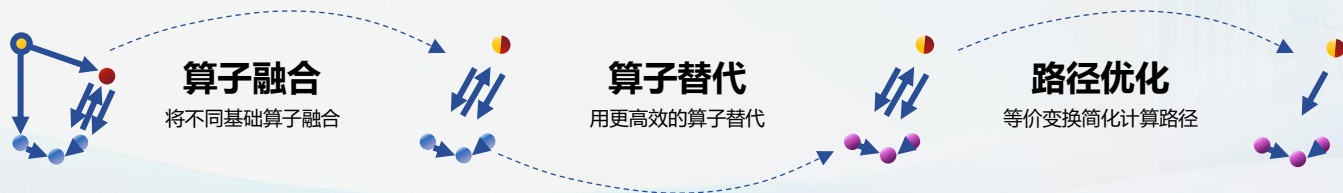
## 基于推理综合性能基准

### AI一体机综合性能

序列长度	并发数量	TTFT 首字延迟 (s)	TPOT 平均每字延迟 (s)
<b>128 / 1000</b>	<b>159</b>	<b>0.92</b>	<b>0.1</b>

# AI编译优化 – 挖掘计算潜力

训练和推理计算开销各降低**15%+**



# 慢节点故障预测与自愈 - 保障算力极致可靠

自动发现故障，百卡秒级、千卡分钟级、万卡十分钟级自愈



# 专家并行通信算法 - 突破通信瓶颈

推理延迟降低3倍以上

## MoE专家模型架构



## MoE专家模型通信方式



## 目标架构自适应的细粒度计算通信协同优化



显著降低通信延迟，提升通信带宽，  
大幅提高推理响应速度

# 联想可持续发展ESG解决方案全景

## ESG外化能力屋顶图

愿景：五年内成为ICT行业内可持续发展（ESG）解决方案的领导者

### ESG咨询服务

战略咨询 | 绿色供应链管理 | 企业碳管理 | 评级提升 | ESG报告 | 生物多样性



气候行动 | 循环经济 | 能源效率

#### 气候行动

联想零碳服务

联想零立方服务

#### 循环经济

IT设备再生服务（回收、处置、翻新）

#### 能源效率

##### 零碳智慧园区

- AI 能耗管理解决方案
- 智慧能源管理解决方案
- 智慧园区运营管理解决方案

##### 低碳智算中心

- 模块化数据中心解决方案
- 液冷数据中心解决方案

社会公益

#### 联想全球学习中心

- 企业人才培养咨询服务
- 企业出海、AI应用与实践、前沿技术相关的战略与经营管理咨询
- 专业共建、联合授课、人才培养基地等

合规治理

#### 联想法务咨询服务

- 合规评估与咨询
- 合规体系建立
- 数字化合规治理平台
- 企业出海&贸易合规
- 应对监管与法律支持

## 企业ESG数字化管理平台（乐循）

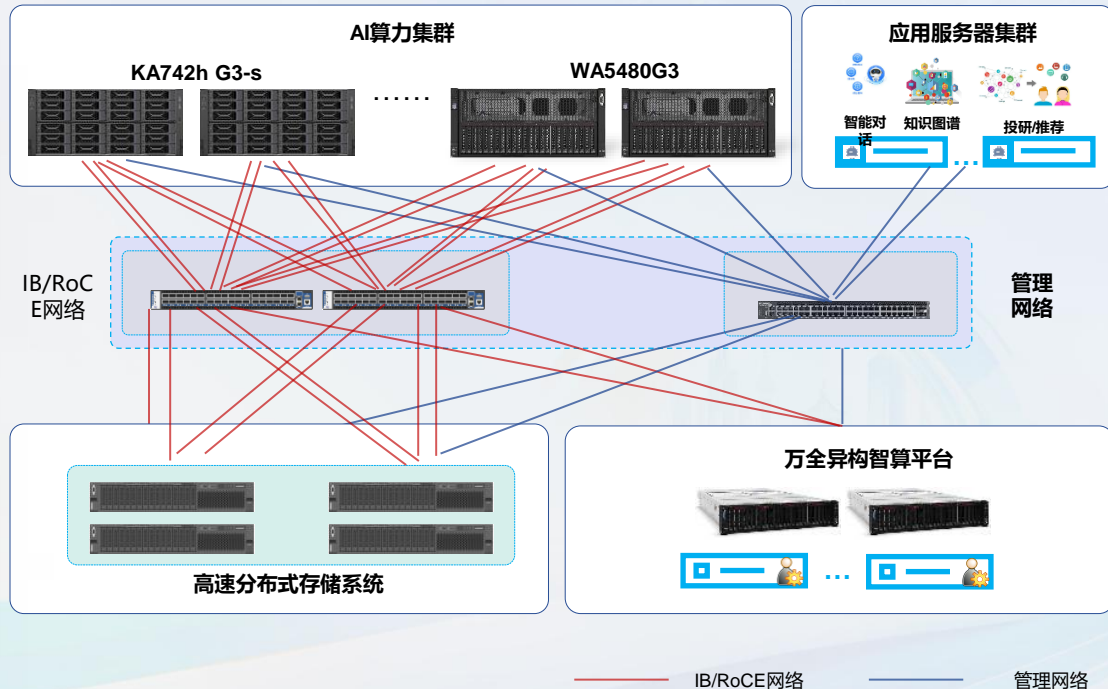
# 某金融机构满血版推理案例

## 项目背景&挑战

- 业务发展要求从“经验驱动”转向“数据+认知驱动”
- 平台建设面临信创和集群管理的多重挑战
- 用于智能投顾、知识库、问答助手等。

## 联想解决方案

- 采用信创与X86混合算力集群
- 集群使用ROCE无损网络互联
- 通过异构管理平台对集群资源进行统一管理和调度。



# 以万全智算，赋能AI跃升