自然语言处理基础

CS2916 大语言模型

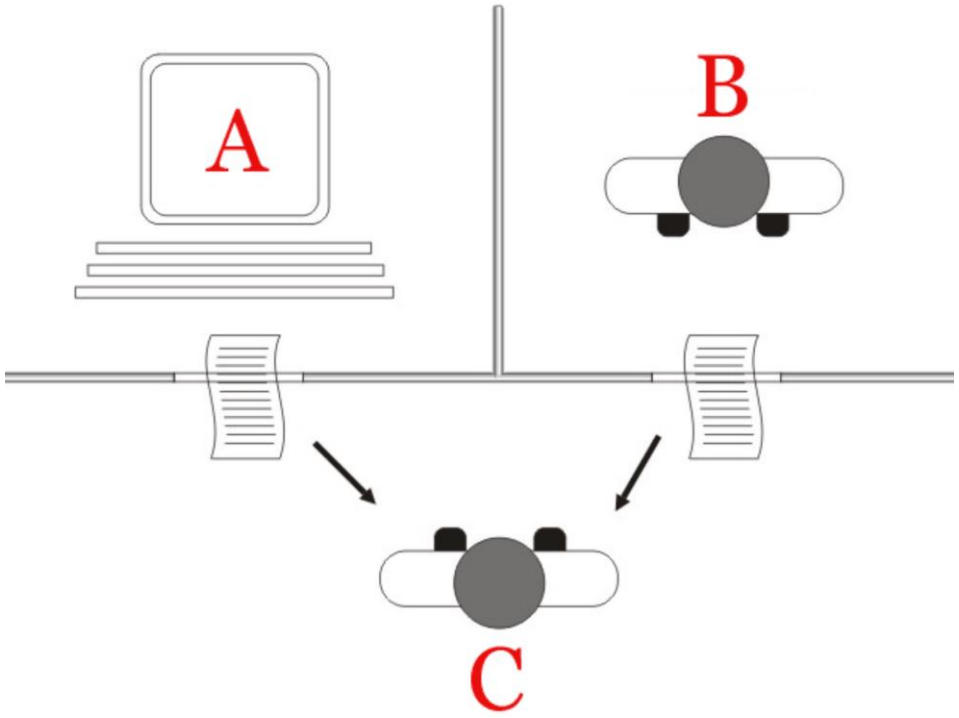
饮水思源 爱国荣校

https://plms.ai/teaching/index.html

图灵测试 (1950)

本质上是一个评估问题，解决方法：Reference-based

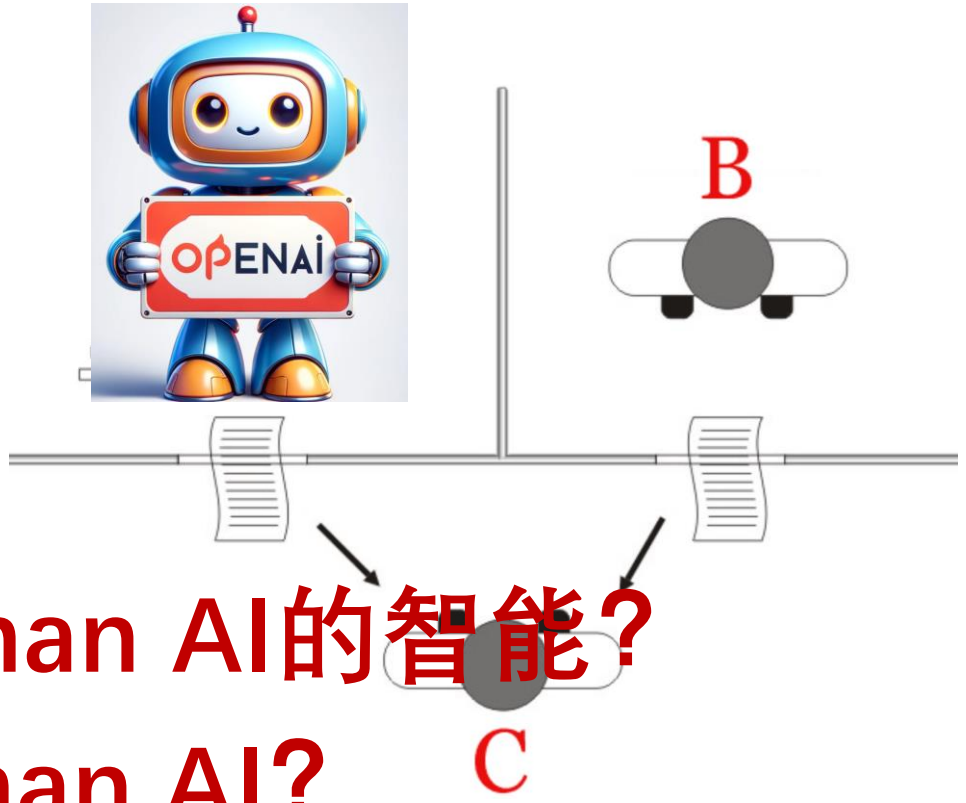
- 目的
  - 检验机器的行为是否类似于人类的智能行为
- 测试方法
  - 能否以人类无法区分的方式思考或表达思考
- 涉及到的技术
  - 自然语言处理、自动推理、计算机视觉、机器人学等



图灵测试 (1950)

本质上是一个评估问题，解决方法：Reference-based

- 目的
  - 检验机器的行为是否类似于人类的智能行为
- 测试方法
  - 能否以人类无法区分的方式思考或表达思考
- 涉及到的技术
  - 自然语言处理、自动推理、计算机视觉、机器人学等



思考：如何评估superhuman AI的智能？

如何训练superhuman AI?

Measuring Progress on Scalable Oversight for Large Language Models， Bowman et al.2022

Scalable Meta-Evaluation of LLMs as Evaluators via Agent Debate， Chern et al.2024

Superalignment: https://openai.com/blog/introducing-superalignment

Weak-to-Strong Generalization: Eliciting Strong Capabilities With Weak Supervision, OpenAI 2024

什么是自然语言处理?

语言学家
刘涌泉

自然语言处理是**人工智能**领域的主要内容，即利用电子计算机 等工具对人类所特有的语言信息（包括口语信息和文字信息）进行各种 加工，并建立各种类型的人-机-人系统。自然语言理解是其核心，其中 包括语音和语符的自动识别以及语音的自动合成。”

Natural language processing (NLP) is an **interdisciplinary** subfield of **computer science** and **linguistics**. It is primarily concerned with giving computers the ability to **support and manipulate human language**.

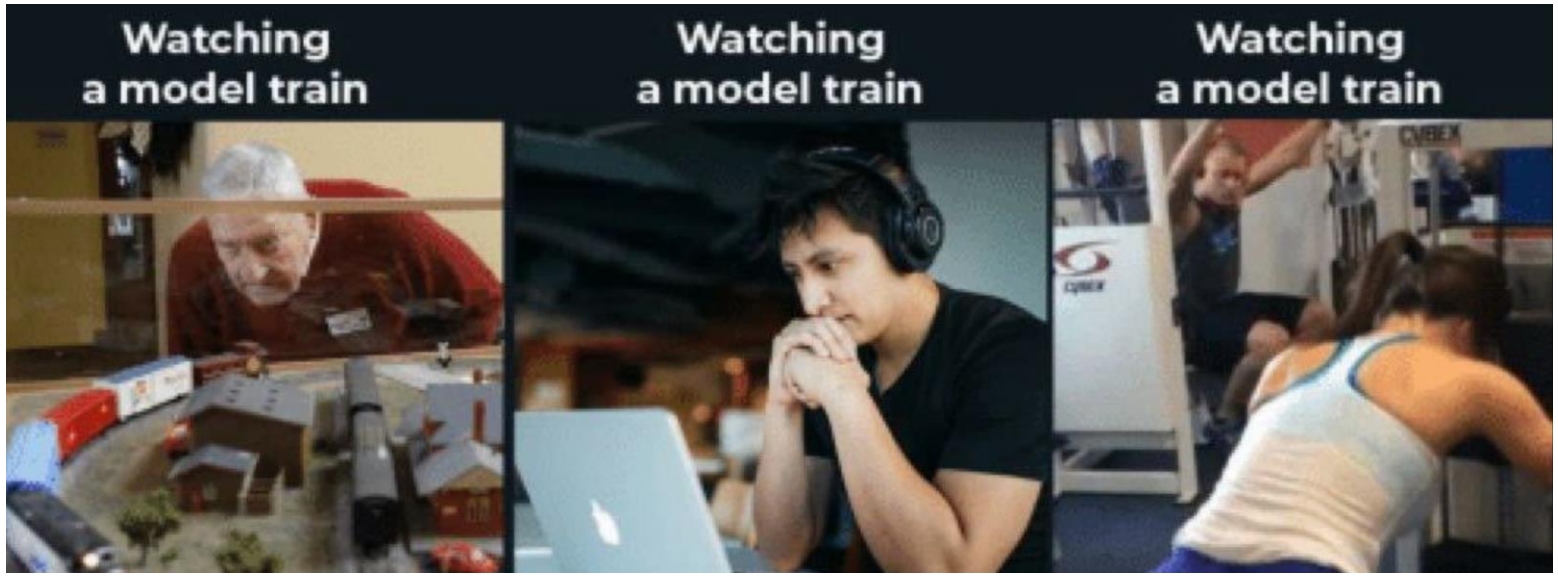
维基百科

宗成庆
老师

我们从事自然语言理解研究的任务也就是研究和探索针对**具体应用**目的的**新方法和新技术**， 使实现系统的性能表现尽量**符合人类理解的标准和要求**。

自然语言处理挑战

多义性



结构复杂性

递归性

“早上点早点早上早点早点吃”

自然语言处理相关书籍/课程

相关书籍

- **Foundations of Statistical Natural Language Processing** Christopher Manning and Hinrich Schütze (1999)
- **An Introduction to Natural Language Processing**, Daniel Jurafsky and James Martin (2008)
- **Neural Network Methods for Natural Language Processing**， Yoav Goldberg
- 自然语言处理综论，宗成庆老师

相关课程

- **交大:** 赵海、俞凯、陈露、林洲汉老师开设了自然语言处理课程
- **Stanford：** CS224n
- **CMU:** CS11-747

自然语言处理相关学术会议

- ACL
  - 成立于1962年，每年一次
  - NLP和计算语言学**最高级别**的会议
  - 在北美、欧洲、亚洲分年会
- EMNLP
  - 发起于1996年，专注于NLP技术的经验方法
  - 随着统计方法和机器学习技术应用广泛而兴起
- NAACL
- TACL （期刊）
  - 每个月1号都可以投稿
  - 审稿周期和ACL相当
- COLING
  - 成立于1965年，两年一次

Categories > Engineering & Computer Science > **Computational Linguistics**

	Publication
1.	Meeting of the Association for Computational Linguistics (ACL)
2.	Conference on Empirical Methods in Natural Language Processing (EMNLP)
3.	Conference of the North American Chapter of the Association for Computational Li Language Technologies (HLT-NAACL)
4.	Transactions of the Association for Computational Linguistics
5.	International Conference on Computational Linguistics (COLING)
6.	International Conference on Language Resources and Evaluation (LREC)
7.	Workshop on Machine Translation
8.	International Workshop on Semantic Evaluation
9.	Conference on Computational Natural Language Learning (CoNLL)
10.	Computer Speech & Language

NLP领域影响力最大的会议（期刊）Top10

根据谷歌学术指标：
https://scholar.google.com/citations?view_op=top_venues&hl=en&vq=eng_computationallinguistics

自然语言处理相关学术会议

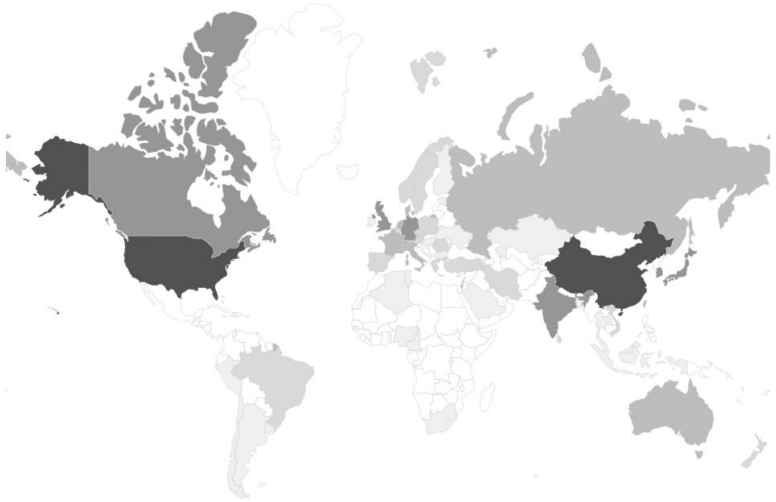
- 其它国际会议：
  - AAAI/IJCAI
  - ICLR/NeurIPS/ICML
- 国内会议
  - CCL
    - 创建于1991年
    - 中国中文信息学会的旗舰会议
  - NLPCC
    - 国际自然语言处理与计算会议

自然语言处理投稿机制

ACL Rolling Review (ARR)

- 背景：整个AI领域发展迅猛，投稿激增，一年一次的会议投稿审稿时间太长，无法满足技术发展与更新迭代
- 发起人：CMU教授Graham Neubig率先提议ARR，两阶段
  - 集中滚动评审
  - 提交投稿至特定会议

Cycle	Submission Date	Author Response	Cycle End
February 2024	Feb 15th	March 24th - 27th	April 15th
April 2024	April 15th		June 15th
June 2024	June 15th		August 15th
August 2024	August 15th		October 15th
October 2024	October 15th		December 15th
December 2024	December 15th		February 15th



https://stats.aclrollingreview.org/

自然语言处理论文集

ACL Anthology

ACL Events

Venue	2023 – 2020	2019 – 2010	2009 – 2000	1999 – 1990	1989 and older
AACL	23 22 20				
ACL	23 22 21 20	19 18 17 16 15 14 13 12 11 10	09 08 07 06 05 04 03 02 01 00	99 98 97 96 95 94 93 92 91 90	89 88 87 86 85 84 83 82 81 80 79
ANLP			00	97 94 92	88 83
CL	23 22 21 20	19 18 17 16 15 14 13 12 11 10	09 08 07 06 05 04 03 02 01 00	99 98 97 96 95 94 93 92 91 90	89 88 87 86 85 84 83 82 81 80 78
CoNLL	23 22 21 20	19 18 17 16 15 14 13 12 11 10	09 08 07 06 05 04 03 02 01 00	99 98 97	
EACL	23 21	17 14 12	09 06 03	99 97 95 93 91	89 87 85 83
EMNLP	23 22 21 20	19 18 17 16 15 14 13 12 11 10	09 08 07 06 05 04 03 02 01 00	99 98 97 96	
Findings	23 22 21 20				
IWSLT	23 22 21 20	19 18 17 16 15 14 13 12 11 10	09 08 07 06 05 04		
NAACL	22 21	19 18 16 15 13 12 10	09 07 06 04 03 01 00		
SemEval	23 22 21 20	19 18 17 16 15 14 13 12 10	07 04 01	98	
*SEM	23 22 21 20	19 18 17 16 15 14 13 12			
TACL	23 22 21 20	19 18 17 16 15 14 13			
WMT	23 22 21 20	19 18 17 16 15 14 13 12 11 10	09 08 07 06		
WS	23 22 21 20	19 18 17 16 15 14 13 12 11 10	09 08 07 06 05 04 03 02 01 00	99 98 97 96 95 94 93 91 90	89 87 85 83 81 79
SIGs	ANN | BIOMED | DAT | DIAL | EDU | EL | FSM | GEN | HAN | HUM | LEX | MEDIA | MOL | MORPHON | MT | NLL | PARSE | REP | SEM | SEMITIC | SLAV | SLPAT | SLT | TYP | UL | UR | WAC				

Non-ACL Events

Venue	2023 – 2020	2019 – 2010	2009 – 2000	1999 – 1990	1989 and older
ALTA	22 21 20	19 18 17 16 15 14 13 12 11 10	09 08 07 06 05 04 03		
AMTA	22 20	18 16 14 12 10	08 06 04 02 00	98 96 94	
CCL	23 22 21 20				
COLING	22 20	18 16 14 12 10	08 06 04 02 00	98 96 94 92 90	88 86 84 82 80 73 69 67
EAMT	23 22 20	18 16 15 14 12 11 10	09 08 06 05 04 03 02 00	99 98 97 96 94 93	
HLT			06 05 04 03 01	94 93 92 91 90	89 86
IJCLCLP	21 20	19 18 17 16 15 14 13 12 11 10	09 08 07 06 05 04 03 02 01 00	99 98 97 96	
IJCNLP	23 22 21	19 17 15 13 11	09 08 05		
JEP/TALN/RECITAL	23 22 21 20	19 18 17 16 15 14 13 12 11 10	09 08 07 06 05 04 03 02 01		
KONVENS	22 21				
LILT		19 18 17 16 15 14			
LREC	22 20	18 16 14 12 10	08 06 04 02 00		
MTSummit	23 21	19 17 15 13 11	09 07 05 03 01	99 97 95 93 91	89 87
MUC				98 95 93 92 91	
NEJLT	22 21				
PACLIC	23 22 21 20	18 17 16 15 14 13 12 11 10	09 08 07 06 05 04 03 01 00	99 98 96 95	
RANLP	23 21	19 17 15 13 11	09		

bib (full) **Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**

pdf bib **Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**
Anna Rogers | Jordan Boyd-Graber | Naoaki Okazaki

pdf bib abs **Program Chairs' Report on Peer Review at ACL 2023**
Anna Rogers | Marzena Karpinska | Jordan Boyd-Graber | Naoaki Okazaki

pdf bib abs **One Cannot Stand for Everyone! Leveraging Multiple User Simulators to train Task-oriented Dialogue Systems**
Yajiao Liu | Xin Jiang | Yichun Yin | Yasheng Wang | Fei Mi | Qun Liu | Xiang Wan | Benyou Wang

pdf bib abs **SafeConv: Explaining and Correcting Conversational Unsafe Behavior**
Mian Zhang | Lifeng Jin | Linfeng Song | Haitao Mi | Wenliang Chen | Dong Yu

pdf bib abs **Detecting and Mitigating Hallucinations in Machine Translation: Model Internal Workings Alone Do Well, Sentence Similarity Even Better**
David Dale | Elena Voita | Loic Barrault | Marta R. Costa-jussà

pdf bib abs **Explainable Recommendation with Personalized Review Retrieval and Aspect Learning**
Hao Cheng | Shuo Wang | Wensheng Lu | Wei Zhang | Mingyang Zhou | Kezhong Lu | Hao Liao

pdf bib abs **Binary and Ternary Natural Language Generation**
Zechun Liu | Barlas Oguz | Aasish Pappu | Yangyang Shi | Raghuraman Krishnamoorthi

pdf bib abs **Span-Selective Linear Attention Transformers for Effective and Robust Schema-Guided Dialogue State Tracking**
Björn Bebensee | Haejun Lee

pdf bib abs **EM Pre-training for Multi-party Dialogue Response Generation**
Yiyang Li | Hai Zhao

pdf bib abs **ACLM: A Selective-Denoising based Generative Data Augmentation Approach for Low-Resource Complex NER**
Sreyan Ghosh | Utkarsh Tyagi | Manan Suri | Sonal Kumar | Ramaneswaran S | Dinesh Manocha

自然语言处理发展：近代史 （1950-2010）

时间	阶段	主要成就
1950年代	早期探索	图灵测试提出
1960-1980年代	规则基础时期	ELIZA等基于规则的对话系统
1990年代	统计方法的革命	隐马尔可夫模型（HMM）等统计模型应用于NLP 支持向量机（SVM）等机器学习算法开始应用
2000年代	机器学习方法的普及	如支持向量机（SVM）、条件随机场（CRF）等在NLP任务中的应用

自然语言处理发展：近代史

时间	阶段	主要成就
1950年代	早期探索	图灵测试提出
1960-1980年代	规则基础时期	ELIZA等基于规则的对话系统
1990年代	统计方法的革命	隐马尔可夫模型（HMM）等统计模型应用于NLP 支持向量机（SVM）等机器学习算法开始应用
2000年代	机器学习方法的普及	如支持向量机（SVM）、条件随机场（CRF）等在NLP任务中的应用

旗舰会议EMNLP会议逐渐形成

Text categorization with support vector machines: Learning with many relevant features

T Joachims Cited by 12924

European conference on machine learning, 1998 • Springer

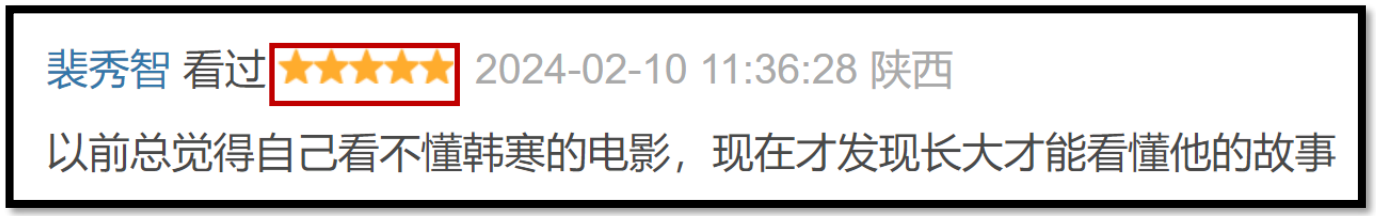
Conditional random fields: Probabilistic models for segmenting and labeling sequence data

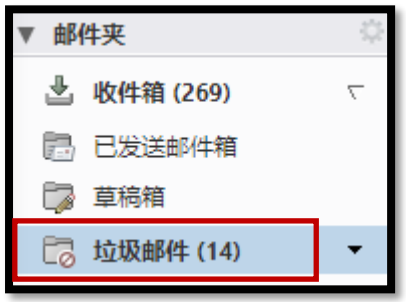
J Lafferty, A McCallum, FCN Pereira Cited by 18432

2001 • repository.upenn.edu

自然语言处理任务：Text to Label

- 任务描述：
  - 输入：一段文本（text）
  - 输出：类别标签（label）
- 具体任务
  - 情感分类
  - 垃圾邮件过滤
- 例子





自然语言处理任务：Text-Span to Label

- 任务描述
  - 输入：一个句子(text)和一个词段(span)
  - 输出：类别标签
- 具体任务
  - 基于“视角”的情感分类
- 例子
  - “这款手机的性能超乎我的期待，速度快，**屏幕**显示效果也非常好。但是，**电池**寿命较短，这让我有些失望。”

自然语言处理任务：Text-Text to Label

- 任务描述：
  - 输入：文本对
  - 输出：类别标签
- 具体任务
  - 自然语言推理 (natural language inference)
- 例子
  - 句子1：*一群孩子在公园里玩耍。*
  - 句子2：
    - *孩子们在室内玩耍*（矛盾）
    - *孩子们在享受户外活动*（蕴含）
    - *公园里正在举行一个生日派对*（中立）

给定两个句子，判断这两者之间的关系是蕴含（entailment）、矛盾（contradiction）还是中立（neutral）

自然语言处理任务：Text to Labels

- 任务描述:
  - 输入：文本
  - 输出：标签序列
- 具体任务
  - 词性标注
  - 命名实体识别
  - 中文分词
- 例子

Sentence (X)	Stefan	Liu	will	graduate	from	Carnegie	Mellon	University
Named Entity	PERSON					Organization		
Tags (Y)	B-PER	E-PER	O	O	O	B-ORG	I-ORG	E-ORG

Sentence (X)	I	ate	two	apples
tags (Y)	PRP	VBD	CD	NNS

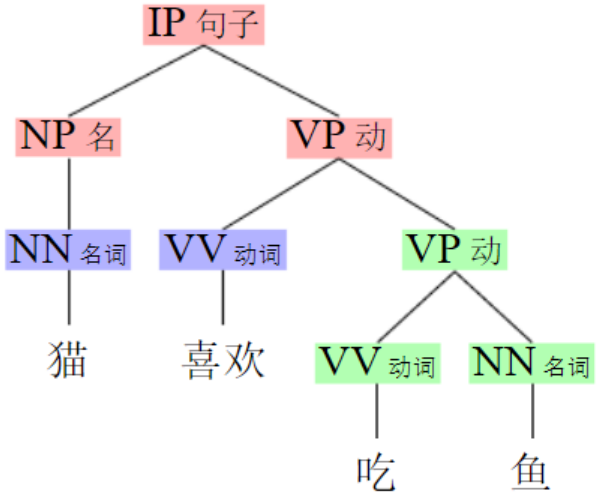
自然语言处理任务：Text to Text

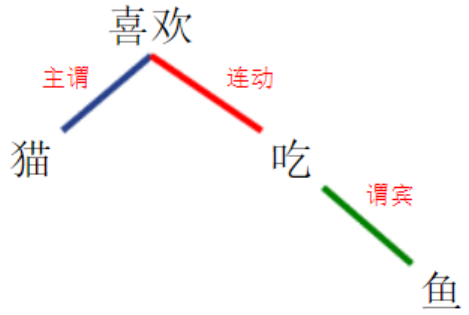
- 任务描述
  - 输入：一段文本
  - 输出：一段文本
- 具体任务
  - 机器翻译
  - 文本摘要
  - 自动对话
- 例子
  - “龙年快乐，万事如意” 的翻译

语言	翻译
英语	Happy Year of the Dragon, may all your wishes come true
西班牙语	Feliz Año del Dragón, que todos tus deseos se hagan realidad
法语	Bonne année du Dragon, que tous vos souhaits se réalisent
日语	龍の年おめでとうございます、万事如意
德语	Glückliches Drachenjahr, mögen alle deine Wünsche in Erfüllung gehen

自然语言处理任务：Text to Tree

- 任务描述
  - 输入：一段文本
  - 输出：树结构的标签
- 具体任务
  - 短语树：描述的是短语的结构功能
  - 依存树：表示了句子中单词和单词之间的依存关系
- 例子

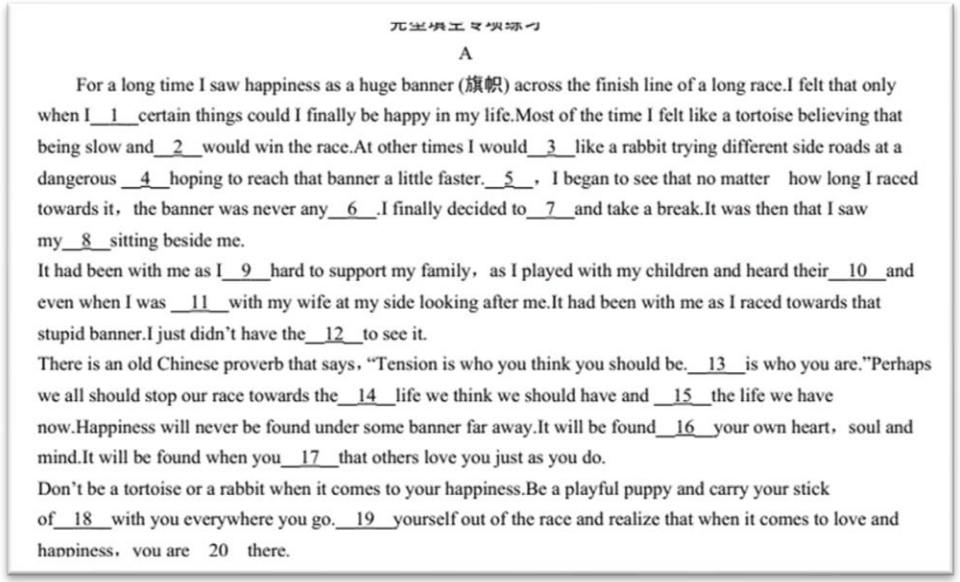




https://opensource.niutrans.com/mtbook/section3-4.html

自然语言处理任务：Word Prediction

- 任务描述：给定一个不完整序列，预测缺失的词元（token)
- 具体任务：
  - 语言模型
  - 词向量学习
- 例子
  - 新年快__



A

For a long time I saw happiness as a huge banner (旗帜) across the finish line of a long race.I felt that only when I__1__certain things could I finally be happy in my life.Most of the time I felt like a tortoise believing that being slow and__2__would win the race.At other times I would__3__like a rabbit trying different side roads at a dangerous __4__hoping to reach that banner a little faster.__5__，I began to see that no matter how long I raced towards it，the banner was never any__6__.I finally decided to__7__and take a break.It was then that I saw my__8__sitting beside me.

It had been with me as I__9__hard to support my family，as I played with my children and heard their__10__and even when I was __11__with my wife at my side looking after me.It had been with me as I raced towards that stupid banner.I just didn't have the__12__to see it.

There is an old Chinese proverb that says，"Tension is who you think you should be.__13__is who you are."Perhaps we all should stop our race towards the__14__life we think we should have and __15__the life we have now.Happiness will never be found under some banner far away.It will be found__16__your own heart，soul and mind.It will be found when you__17__that others love you just as you do.

Don't be a tortoise or a rabbit when it comes to your happiness.Be a playful puppy and carry your stick of__18__with you everywhere you go.__19__yourself out of the race and realize that when it comes to love and happiness，you are 20 there.

自然语言处理中一些重要的概念：Prediction Task

- Text Classification (text -> label): `task-textclass`
- Text Pair Classification (two texts -> label: `task-textpair`
- Sequence Labeling (text -> one label per token): `task-seqlab`
- Extractive Summarization (text -> subset of text): `task-extractive` (implies `text-seqlab`)
- Span Labeling (text -> labels on spans): `task-spanlab`
- Language Modeling (predict probability of text): `task-lm`
- Conditioned Language Modeling (some input -> text): `task-condlm` (implies `task-lm`)
- Sequence-to-sequence Tasks (text -> text, including MT): `task-seq2seq` (implies `task-condlm`)
- Cloze-style Prediction, Masked Language Modeling (right and left context -> word): `task-cloze`
- Context Prediction (as in word2vec) (word -> right and left context): `task-context`
- Relation Prediction (text -> graph of relations between words, including dependency parsing): `task-relation`
- Tree Prediction (text -> tree, including syntactic and some semantic semantic parsing): `task-tree`
- Graph Prediction (text -> graph not necessarily between nodes): `task-graph`
- Lexicon Induction/Embedding Alignment (text/embeddings -> bi- or multi-lingual lexicon): `task-lexicon`
- Word Alignment (parallel text -> alignment between words): `task-alignment`

https://github.com/neulab/nn4nlp-concepts/blob/master/concepts.md

自然语言处理中一些重要的概念： Optimization/Learning

Optimizers and Optimization Techniques

- Mini-batch SGD: `optim-sgd`
- Adam: `optim-adam` (implies `optim-sgd`)
- Adagrad: `optim-adagrad` (implies `optim-sgd`)
- Adadelta: `optim-adadelta` (implies `optim-sgd`)
- Adam with Specialized Transformer Learning Rate ("Noam" Schedule): `optim-noam` (implies `optim-adam`)
- SGD with Momentum: `optim-momentum` (implies `optim-sgd`)
- AMS: `optim-amsgrad` (implies `optim-sgd`)
- Projection / Projected Gradient Descent: `optim-projection` (implies `optim-sgd`)

Initialization

- Glorot/Xavier Initialization: `init-glorot`
- He Initialization: `init-he`

Regularization

- Dropout: `reg-dropout`
- Word Dropout: `reg-worddropout` (implies `reg-dropout`)
- Norm (L1/L2) Regularization: `reg-norm`
- Early Stopping: `reg-stopping`
- Patience: `reg-patience` (implies `reg-stopping`)
- Weight Decay: `reg-decay`
- Label Smoothing: `reg-labelsmooth`

Loss Functions (other than cross-entropy)

- Canonical Correlation Analysis (CCA): `loss-cca`
- Singular Value Decomposition (SVD): `loss-svd`
- Margin-based Loss Functions: `loss-margin`
- Contrastive Loss: `loss-cons`
- Noise Contrastive Estimation (NCE): `loss-nce` (implies `loss-cons`)
- Triplet Loss: `loss-triplet` (implies `loss-cons`)

Training Paradigms

- Multi-task Learning (MTL): `train-mtl`
- Multi-lingual Learning (MLL): `train-mll` (implies `train-mtl`)
- Transfer Learning: `train-transfer`
- Active Learning: `train-active`
- Data Augmentation: `train-augment`
- Curriculum Learning: `train-curriculum`
- Parallel Training: `train-parallel`

自然语言处理中一些重要的概念: Neural Arcs

Activation Functions

- Hyperbolic Tangent (tanh): `activ-tanh`
- Rectified Linear Units (RelU): `activ-relu`

Pooling Operations

- Max Pooling: `pool-max`
- Mean Pooling: `pool-mean`
- k-Max Pooling: `pool-kmax`

Recurrent Architectures

- Recurrent Neural Network (RNN): `arch-rnn`
- Bi-directional Recurrent Neural Network (Bi-RNN): `arch-birnn` (implies `arch-rnn`)
- Long Short-term Memory (LSTM): `arch-lstm` (implies `arch-rnn`)
- Bi-directional Long Short-term Memory (LSTM): `arch-bilstm` (implies `arch-birnn`, `arch-lstm`)
- Gated Recurrent Units (GRU): `arch-gru` (implies `arch-rnn`)
- Bi-directional Gated Recurrent Units (GRU): `arch-bigru` (implies `arch-birnn`, `arch-gru`)

Other Sequential/Structured Architectures

- Bag-of-words, Bag-of-embeddings, Continuous Bag-of-words (BOW): `arch-bow`
- Convolutional Neural Networks (CNN): `arch-cnn`
- Attention: `arch-att`
- Self Attention: `arch-selfatt` (implies `arch-att`)
- Recursive Neural Network (RecNN): `arch-recnn`
- Tree-structured Long Short-term Memory (TreeLSTM): `arch-treelstm` (implies `arch-recnn`)
- Graph Neural Network (GNN): `arch-gnn`
- Graph Convolutional Neural Network (GCNN): `arch-gcnn` (implies `arch-gnn`)

Architectural Techniques

- Residual Connections (ResNet): `arch-residual`
- Gating Connections, Highway Connections: `arch-gating`
- Memory: `arch-memo`
- Copy Mechanism: `arch-copy`
- Bilinear, Biaffine Models: `arch-bilinear`
- Coverage Vectors/Penalties: `arch-coverage`
- Subword Units: `arch-subword`
- Energy-based, Globally-normalized Mdels: `arch-energy`

Standard Composite Architectures

- Transformer: `arch-transformer` (implies `arch-selfatt`, `arch-residual`, `arch-layernorm`, `optim-noam`)

自然语言处理中一些重要的概念: 其它

Composite Pre-trained Embedding Techniques

- word2vec: `pre-word2vec` (implies `arch-cbow`, `task-cloze`, `task-context`)
- fasttext: `pre-fasttext` (implies `arch-cbow`, `arch-subword`, `task-cloze`, `task-context`)
- GloVe: `pre-glove`
- Paragraph Vector (ParaVec): `pre-paravec`
- Skip-thought: `pre-skipthought` (implies `arch-lstm`, `task-seq2seq`)
- ELMo: `pre-elmo` (implies `arch-bilstm`, `task-lm`)
- BERT: `pre-bert` (implies `arch-transformer`, `task-cloze`, `task-textpair`)
- Universal Sentence Encoder (USE): `pre-use` (implies `arch-transformer`, `task-seq2seq`)

Structured Models/Algorithms

- Hidden Markov Models (HMM): `struct-hmm`
- Conditional Random Fields (CRF): `struct-crf`
- Context-free Grammar (CFG): `struct-cfg`
- Combinatorial Categorical Grammar (CCG): `struct-ccg`

Relaxation/Training Methods for Non-differentiable Functions

- Complete Enumeration: `nondif-enum`
- Straight-through Estimator: `nondif-straightthrough`
- Gumbel Softmax: `nondif-gumbelsoftmax`
- Minimum Risk Training: `nondif-minrisk`
- REINFORCE: `nondif-reinforce`

Adversarial Methods

- Generative Adversarial Networks (GAN): `adv-gan`
- Adversarial Feature Learning: `adv-feat`
- Adversarial Examples: `adv-examp`

Latent Variable Models

- Variational Auto-encoder (VAE): `latent-vae`
- Topic Model: `latent-topic`

Meta Learning

- Meta-learning Initialization: `meta-init`
- Meta-learning Optimizers: `meta-optim`
- Meta-learning Loss functions: `meta-loss`
- Neural Architecture Search: `meta-arch`

自然语言处理发展：现代史 （2010-现在）

- **Feature Engineering**
- Architecture Engineering
- Objective Engineering
- Prompt Engineering

- **Paradigm**： Fully Supervised Learning (Non-neural Network)
- **Date**： Before 2013
- **Characteristic**： Traditional machine learning model is mainly used, which requires manual feature definition of input text
- **Typical Work**：
  - CRF (Conditional Random Field)

自然语言处理发展：现代史 （2010-现在）

- Feature Engineering
- **Architecture Engineering**
- Objective Engineering
- Prompt Engineering

- **Paradigm**： Fully Supervised Learning (Neural Network)
- **Date**： 2013 - 2018
- **Characteristic**：
  - Rely on neural networks
  - Do not need to manually define features, but should explore the network structure （e.g.：LSTM v.s CNN）
- **Typical Work**：
  - CNN for Text Classification

自然语言处理发展：现代史 （2010-现在）

- Feature Engineering
- Architecture Engineering
- **Objective Engineering**
- Prompt Engineering

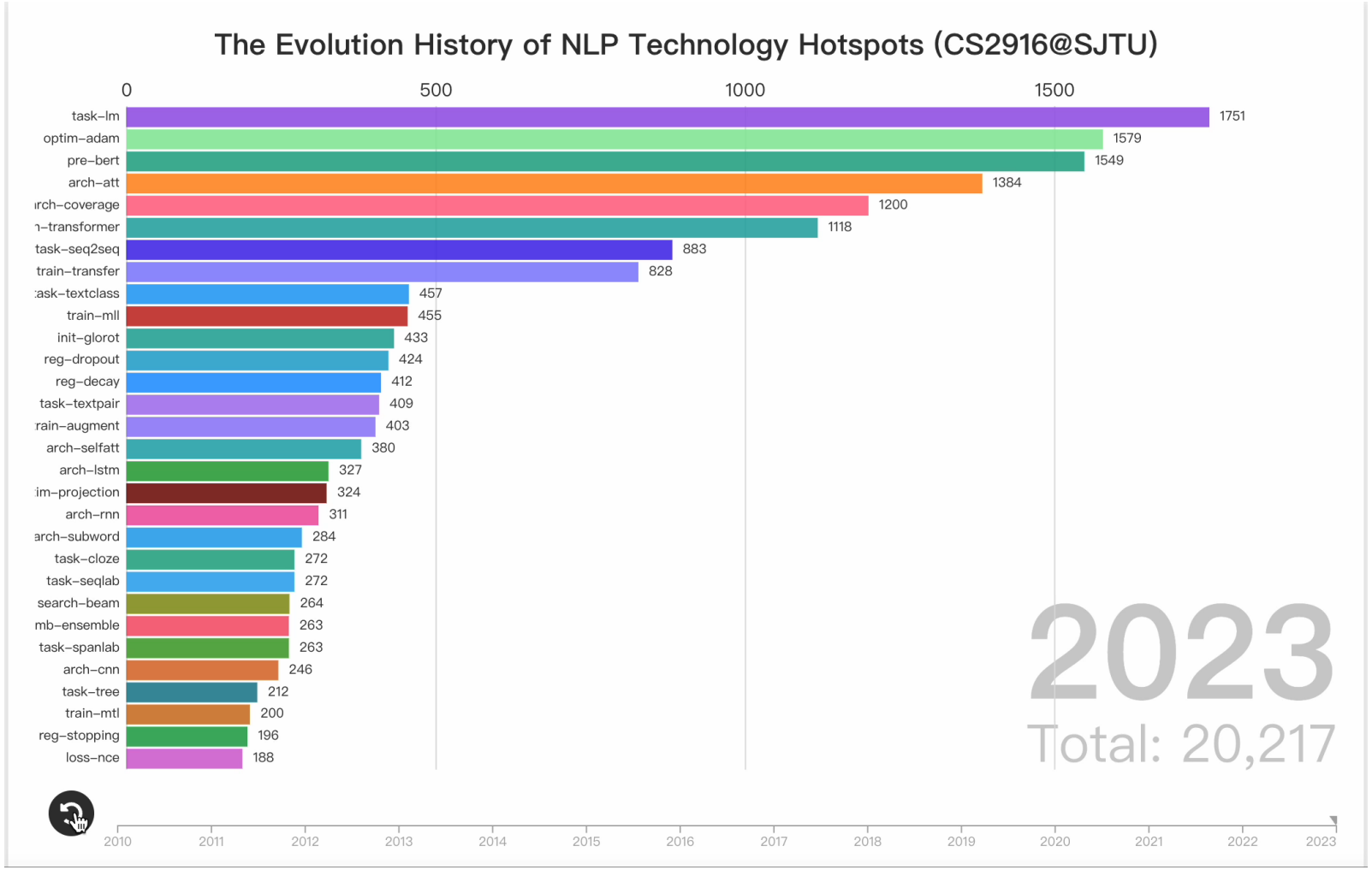
- **Paradigm**： Pre-train, Fine-tune
- **Date**： 2018-Now
- **Characteristic**：
  - context-dependent PLMs
  - Need to pay attention to the definition and selection of objective functions
- **Typical Work**： BERT

自然语言处理发展：现代史 （2010-现在）

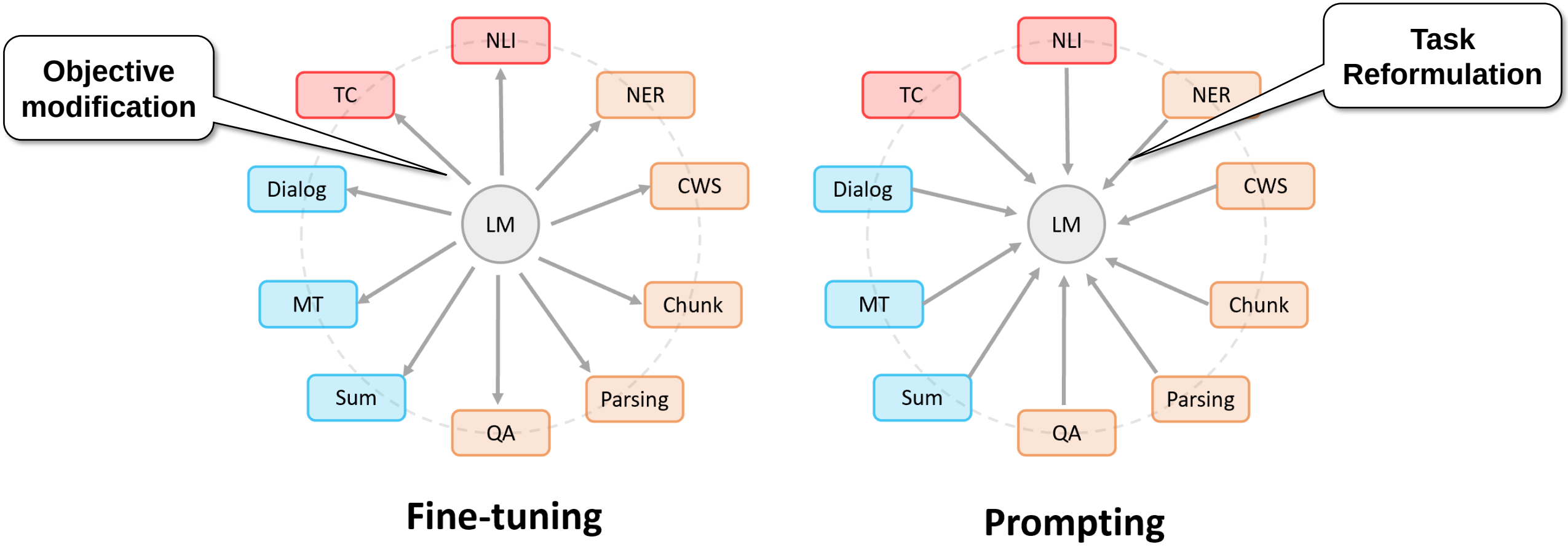
- Feature Engineering
- Architecture Engineering
- Objective Engineering
- **Prompt Engineering**

- **Paradigm**： Pre-train, Prompt, Predict
- **Date**： 2019-Now
- **Characteristic**：
  - NLP tasks are modeled entirely by relying on PLMs
  - More efforts on prompt design
- **Typical Work:** GPT3

NLP技术重要概念的变迁史

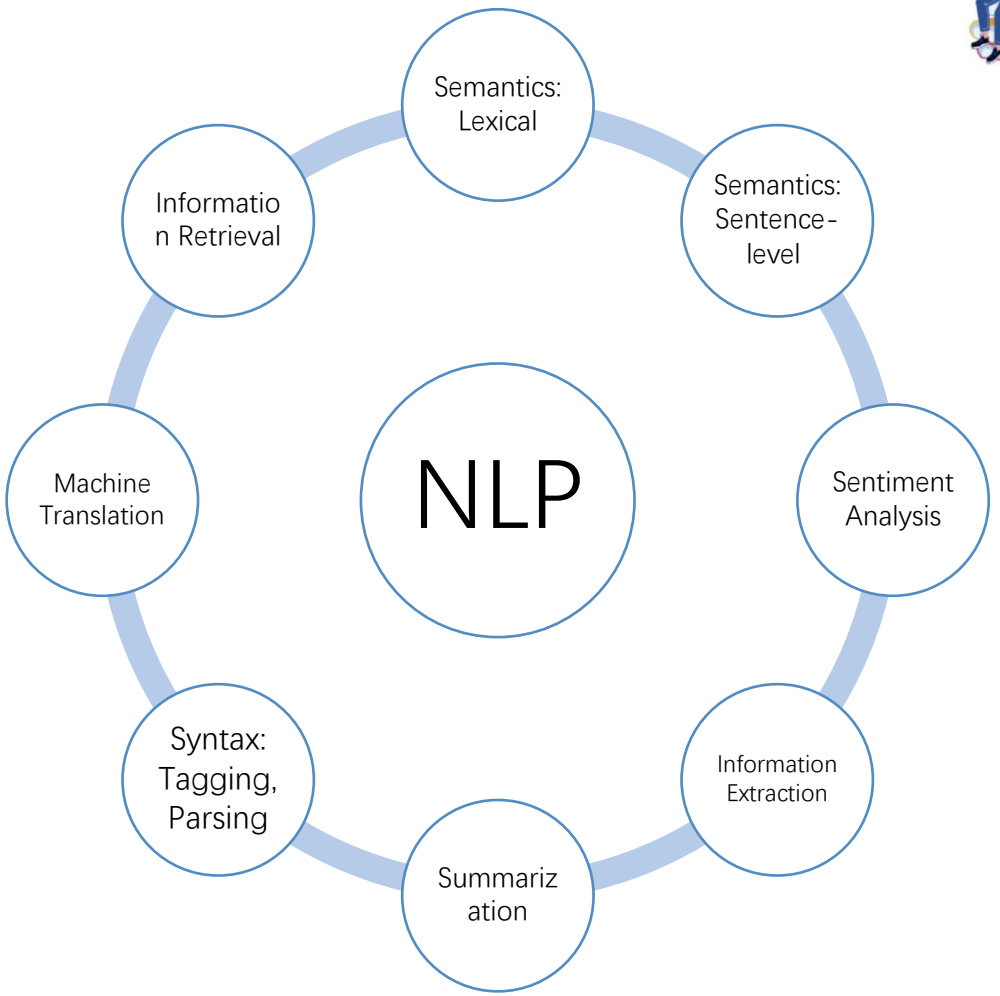


任务的“大一统”

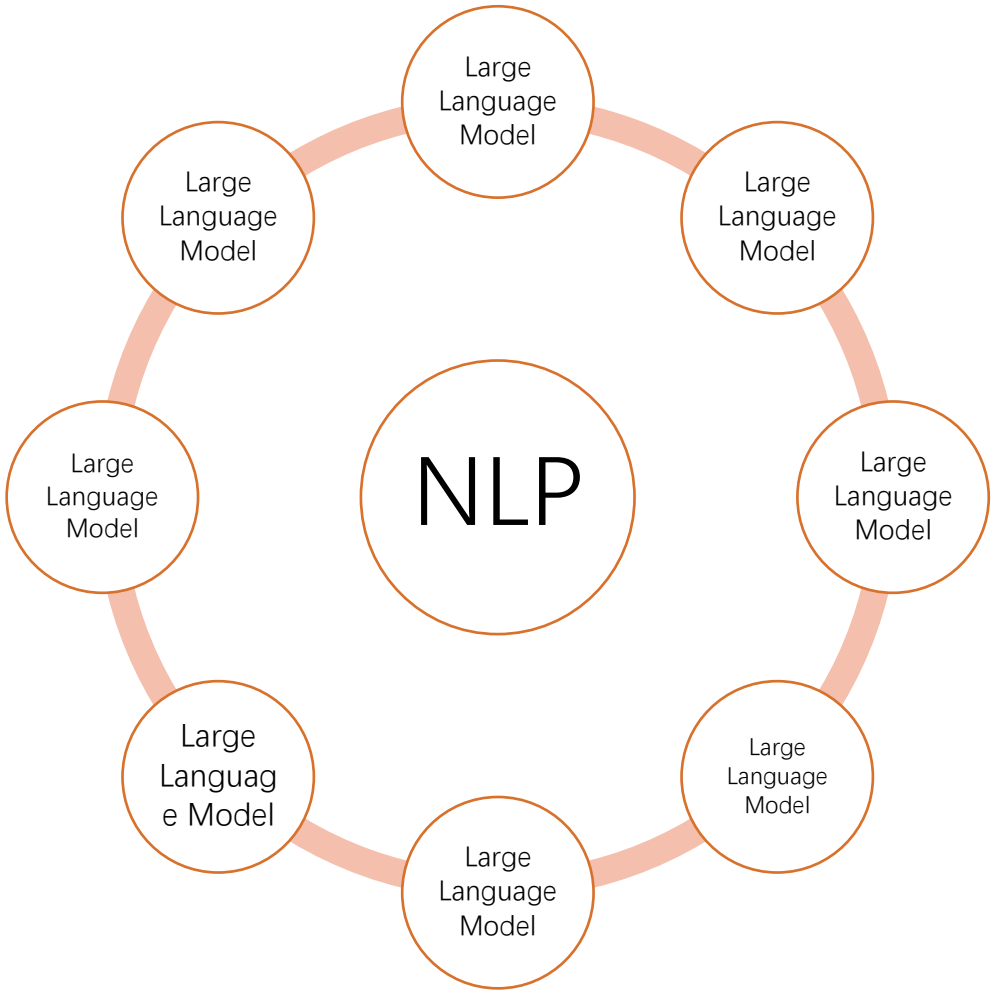

Objective modification
NLI
TC
NER
Dialog
CWS
LM
MT
Chunk
Sum
Parsing
QA
Fine-tuning
Task Reformulation
NLI
TC
NER
Dialog
CWS
LM
MT
Chunk
Sum
Parsing
QA
Prompting

技术范式变革 推动 科研范式改变

过去：任务划分



现在：走向大模型



技术范式变革 推动 科研范式改变

Chelsea Finn @chelseabfinn · Apr 1

In light of tremendous AI advances & recent calls for pauses on AGI research, I've decided to pivot our lab's direction.

Going forward, we'll focus entirely on certain real-world applications (see thread 🧵 for details). It was a hard decision but I'm excited for our future work

46 90 1,061 612.6K

Sasha Rush
@srush_nlp

Irresponsible thought:
ICLM - International Conference on Language Modeling. Would be ICLR-like with topics on applications, efficiency, evaluation, open-source, interpretability, "safety", architectures.

Denny Zhou
@denny_zhou

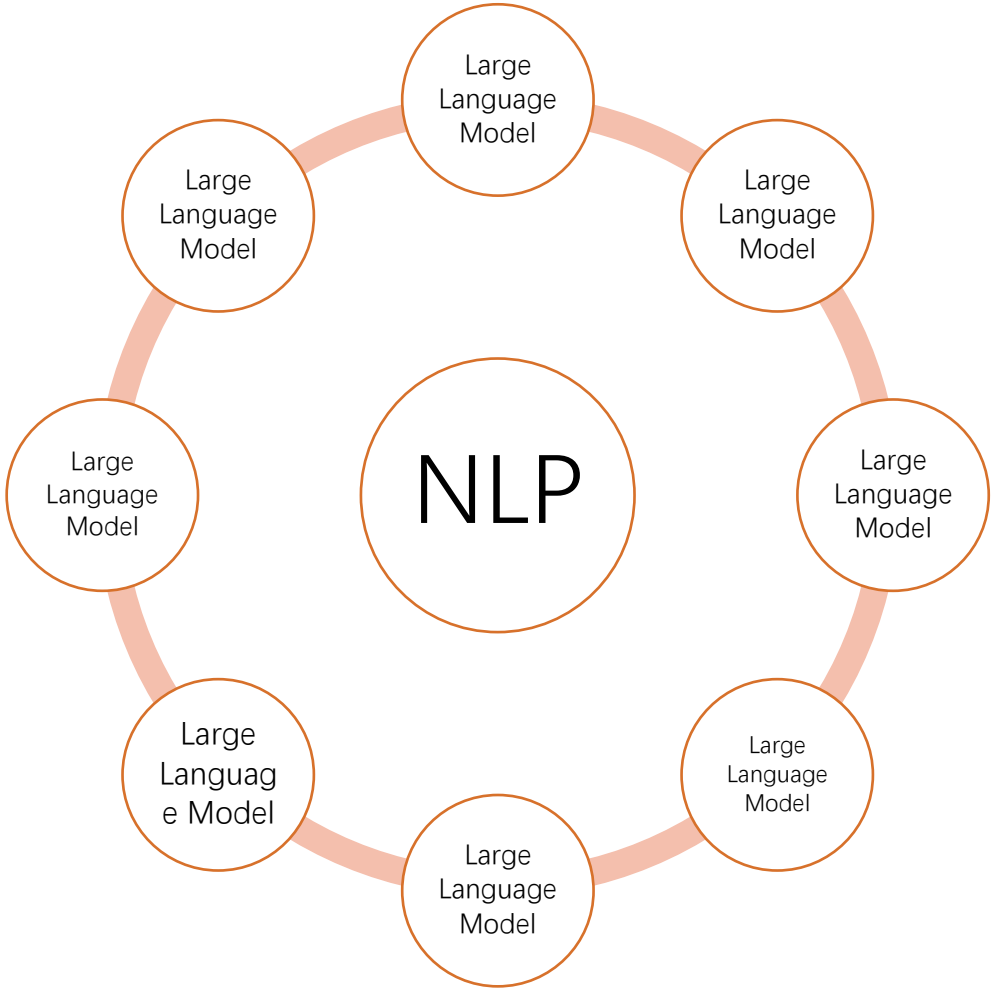
Maybe time to initiate a new conference dedicated to LLMs, reminiscent of how ICLR emerged for DL years ago. This could also help reduce submissions to NeurIPS and ICLR. Any thoughts?

Syntax: Tagging, Parsing

Summarization

Information Extraction

现在：走向大模型



技术范式变革 推动 科研范式改变

ACL

*** Research Area**

Research Areas / Tracks. Select the most relevant research area / track for your paper. This

- ○ Computational Social Science and Cultural Analytics
- ○ Dialogue and Interactive Systems
- ○ Discourse and Pragmatics
- ○ Efficient/Low-Resource Methods for NLP
- ○ Ethics, Bias, and Fairness
- ◉ Generation
- ○ Information Extraction
- ○ Information Retrieval and Text Mining
- ○ Interpretability and Analysis of Models for NLP
- ○ Linguistic theories, Cognitive Modeling and Psycholinguistics
- ○ Machine Learning for NLP
- ○ Machine Translation
- ○ Multilinguality and Language Diversity
- ○ Multimodality and Language Grounding to Vision, Robotics and Beyond
- ○ Phonology, Morphology and Word Segmentation
- ○ Question Answering
- ○ Resources and Evaluation
- ○ Semantics: Lexical
- ○ Semantics: Sentence-level Semantics, Textual Inference and Other areas
- ○ Sentiment Analysis, Stylistic Analysis, and Argument Mining
- ○ Speech recognition, text-to-speech and spoken language understanding
- ○ Summarization
- ○ Syntax: Tagging, Chunking and Parsing / ML
- ○ NLP Applications
- ○ Special Theme (conference specific)

COLM

Call for Papers:

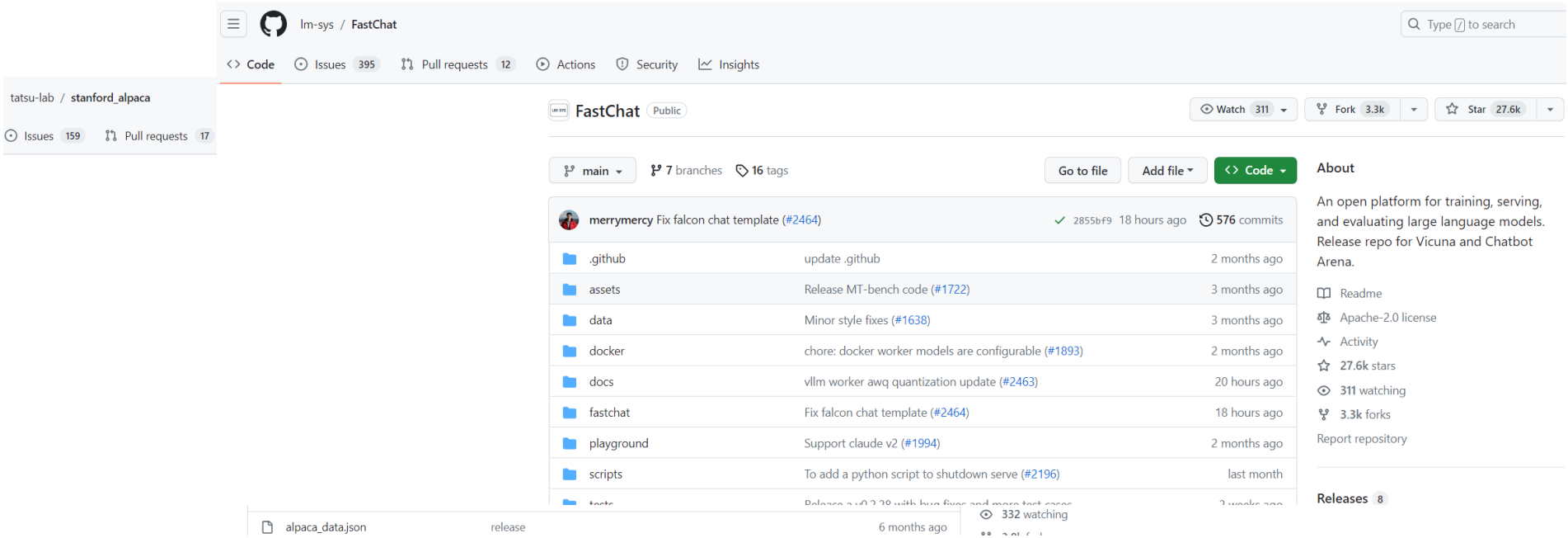
We consider a broad range of subject areas focused on language modeling for t
"language model" in the broadest way. A non-exhaustive list of topics of interes

1. All about **alignment**: fine-tuning, instruction-tuning, reinforcement learnir context alignment
2. All about **data**: pre-training data, alignment data, and synthetic data --- via generation
3. All about **evaluation**: benchmarks, simulation environments, scalable over and/or machine evaluation
4. All about **societal implications**: bias, equity, misuse, jobs, climate change, a
5. All about **safety**: security, privacy, misinformation, adversarial attacks and
6. **Science of LMs**: scaling laws, fundamental limitations, emergent capabiliti training dynamics, grokking, learning theory for LMs
7. **Compute efficient LMs**: distillation, compression, quantization, sample efl
8. **Engineering for large LMs**: distributed training and inference on different instability
9. **Learning algorithms** for LMs: learning, *un*learning, meta learning, model m
10. **Inference algorithms** for LMs: decoding algorithms, reasoning algorithms,
11. **Human mind, brain, philosophy, laws and LMs:** cognitive science, neuros philosophical, or legal perspectives on LMs
12. LMs for **everyone**: multi-linguality, low-resource languages, vernacular lar
13. LMs and **the world**: factuality, retrieval-augmented LMs, knowledge mode social norms, pragmatics, and world models
14. LMs and **embodiment**: perception, action, robotics, and multimodality

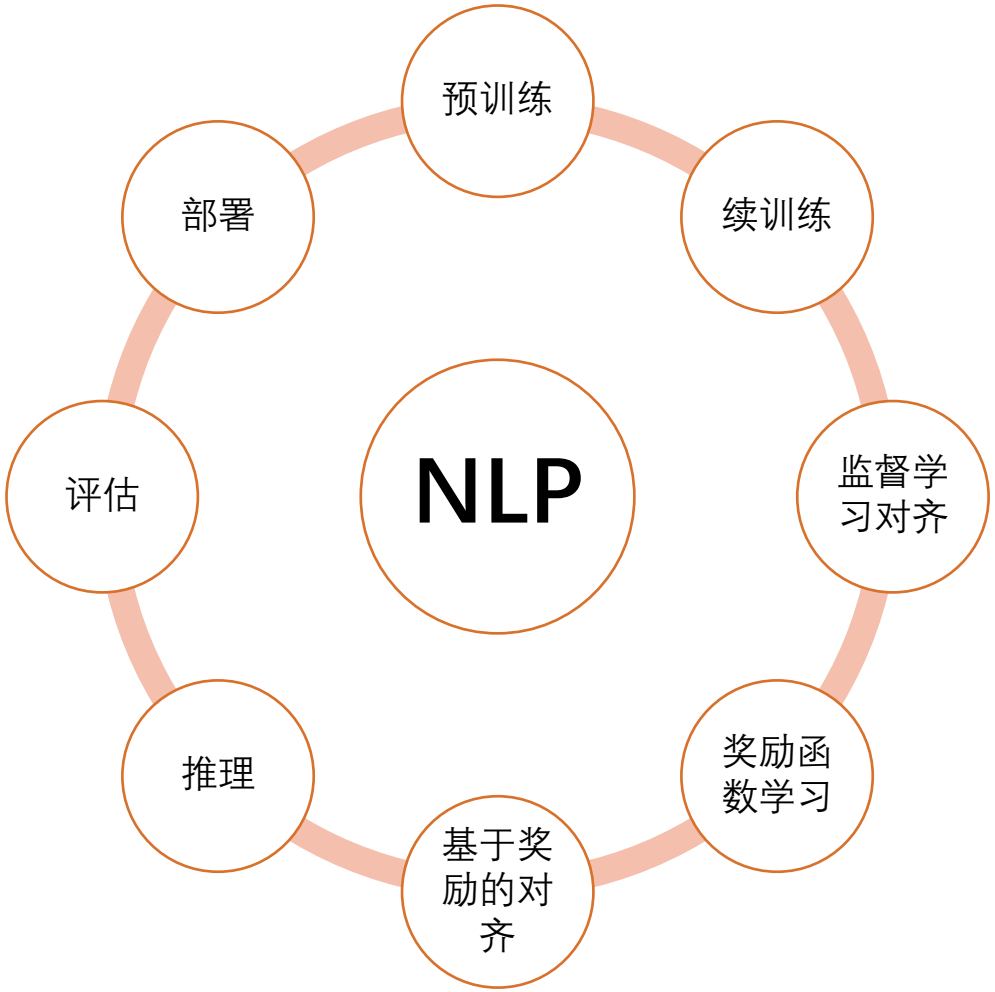
技术范式变革 推动 科研范式改变

- 10年前: ACL/EMNLP/ICML
- 5年前：ArXiv
- 现在：Github/Twitter/最新博文

最新技术获取源发生变化

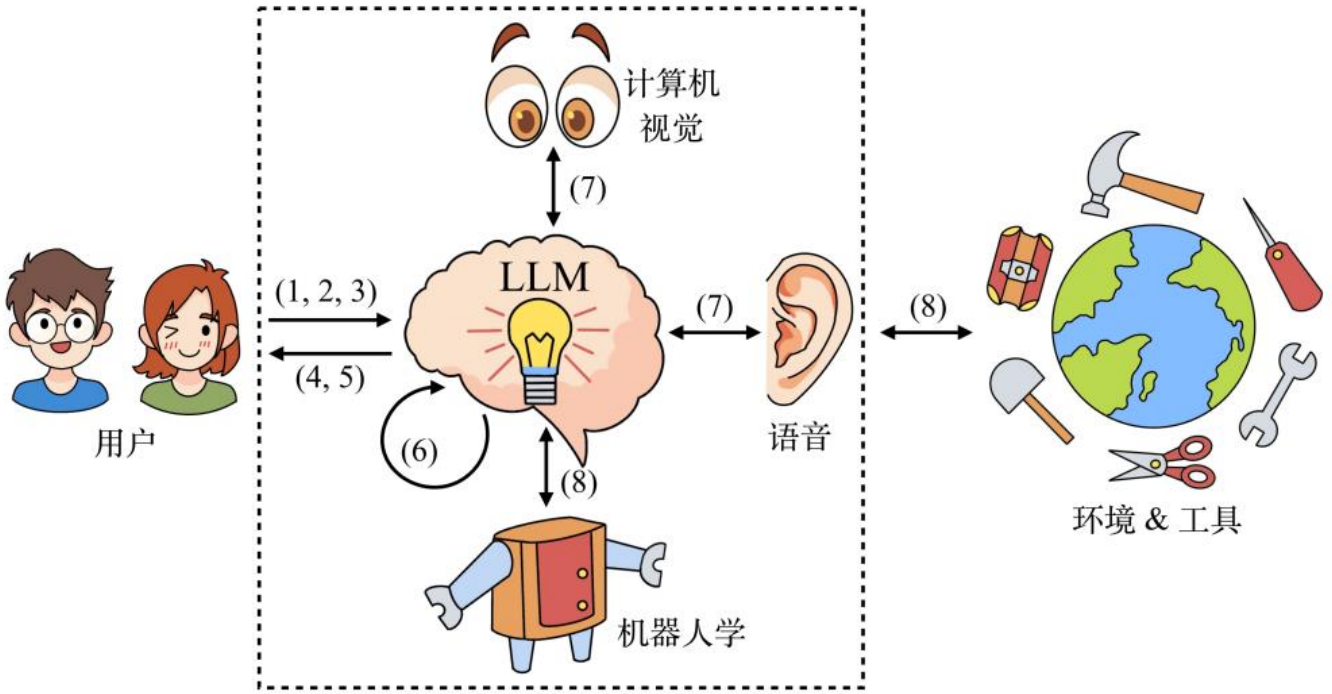


大模型在做什么?



技术栈视角（LLMOps）

大模型在做什么?

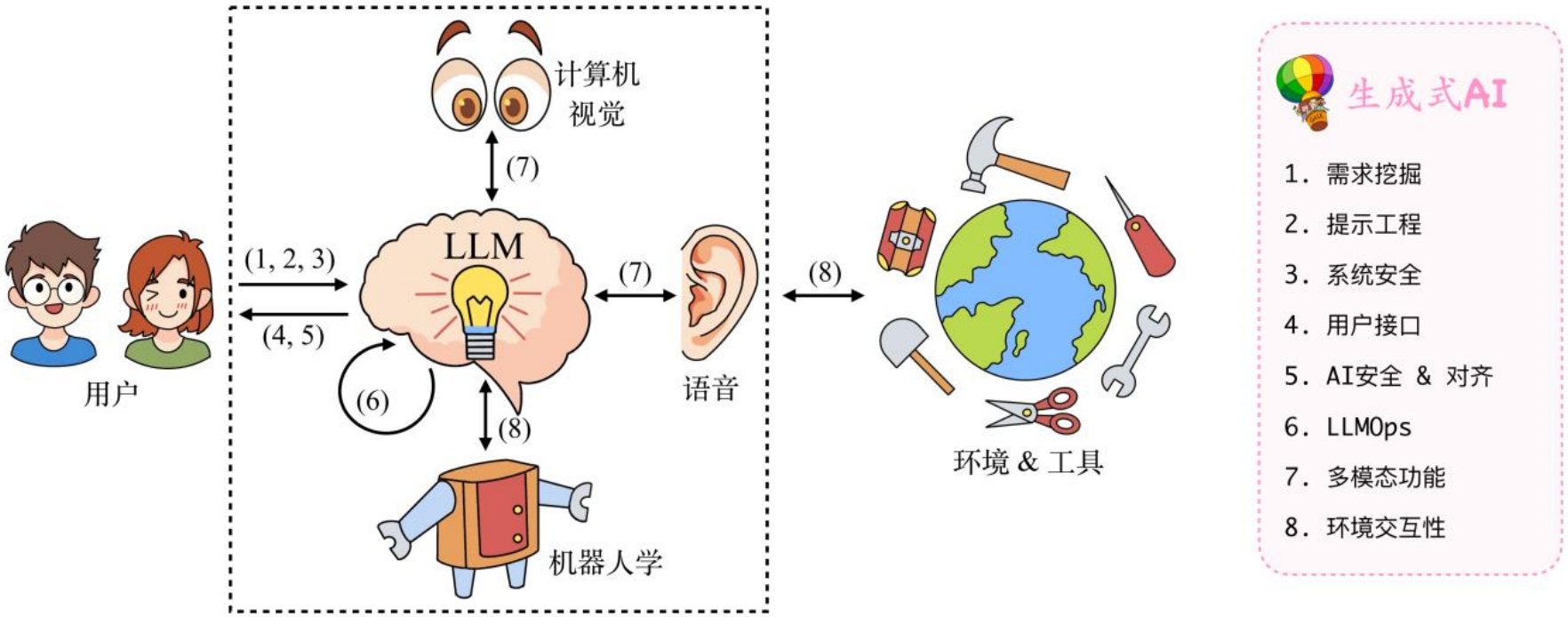


生成式AI

1. 需求挖掘
2. 提示工程
3. 系统安全
4. 用户接口
5. AI安全 & 对齐
6. LLMOps
7. 多模态功能
8. 环境交互性

全景技术栈

新的时代如何展开科研?



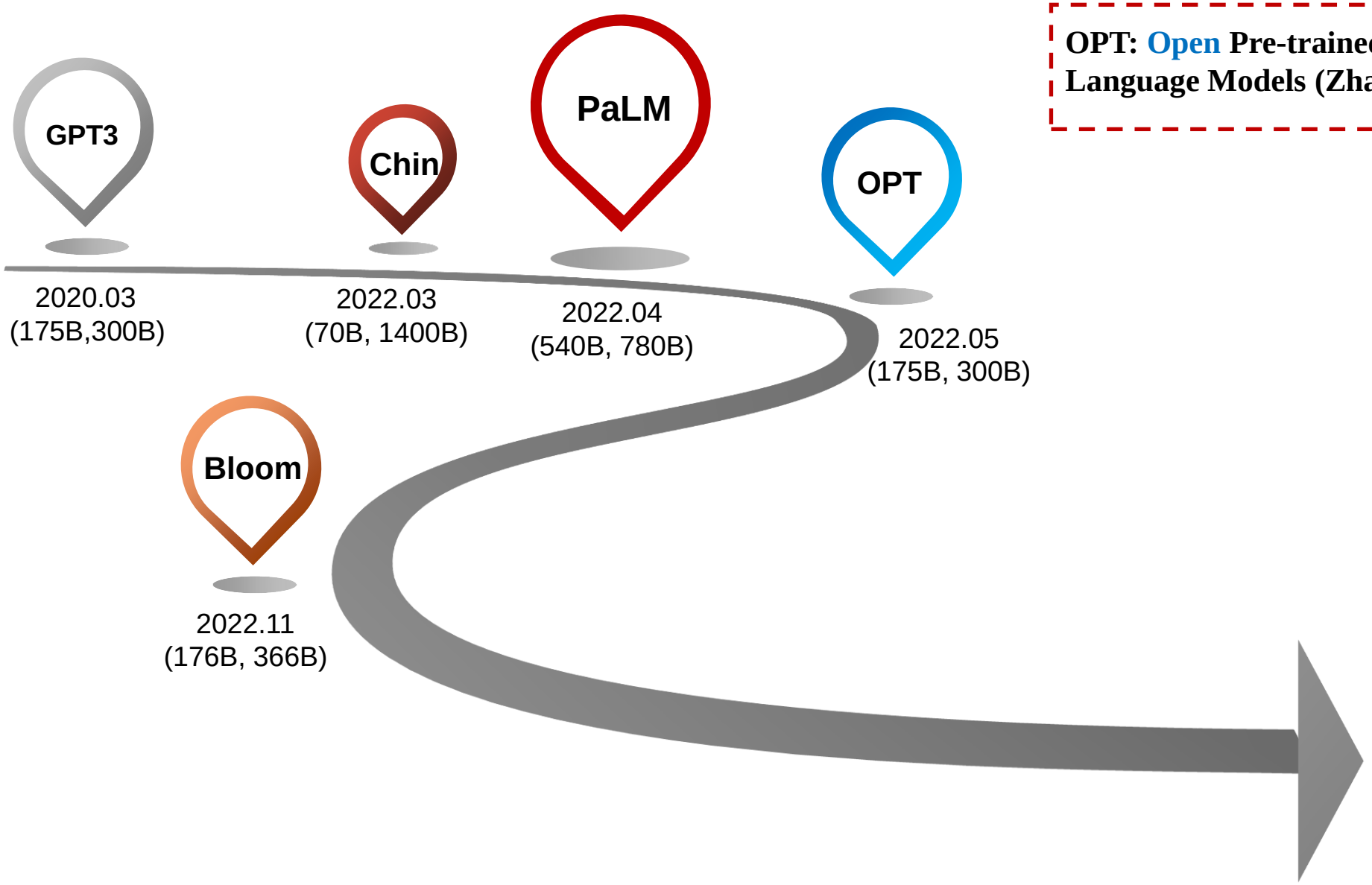
生成式AI

1. 需求挖掘
2. 提示工程
3. 系统安全
4. 用户接口
5. AI安全 & 对齐
6. LLMOps
7. 多模态功能
8. 环境交互性

生成式人工智能时代，研究机构可以研究的问题并没有减少，更多的只是内容上的更新，这也就要求学者敢于定义新任务，新场景，快速试错，并提出可能的解决方案

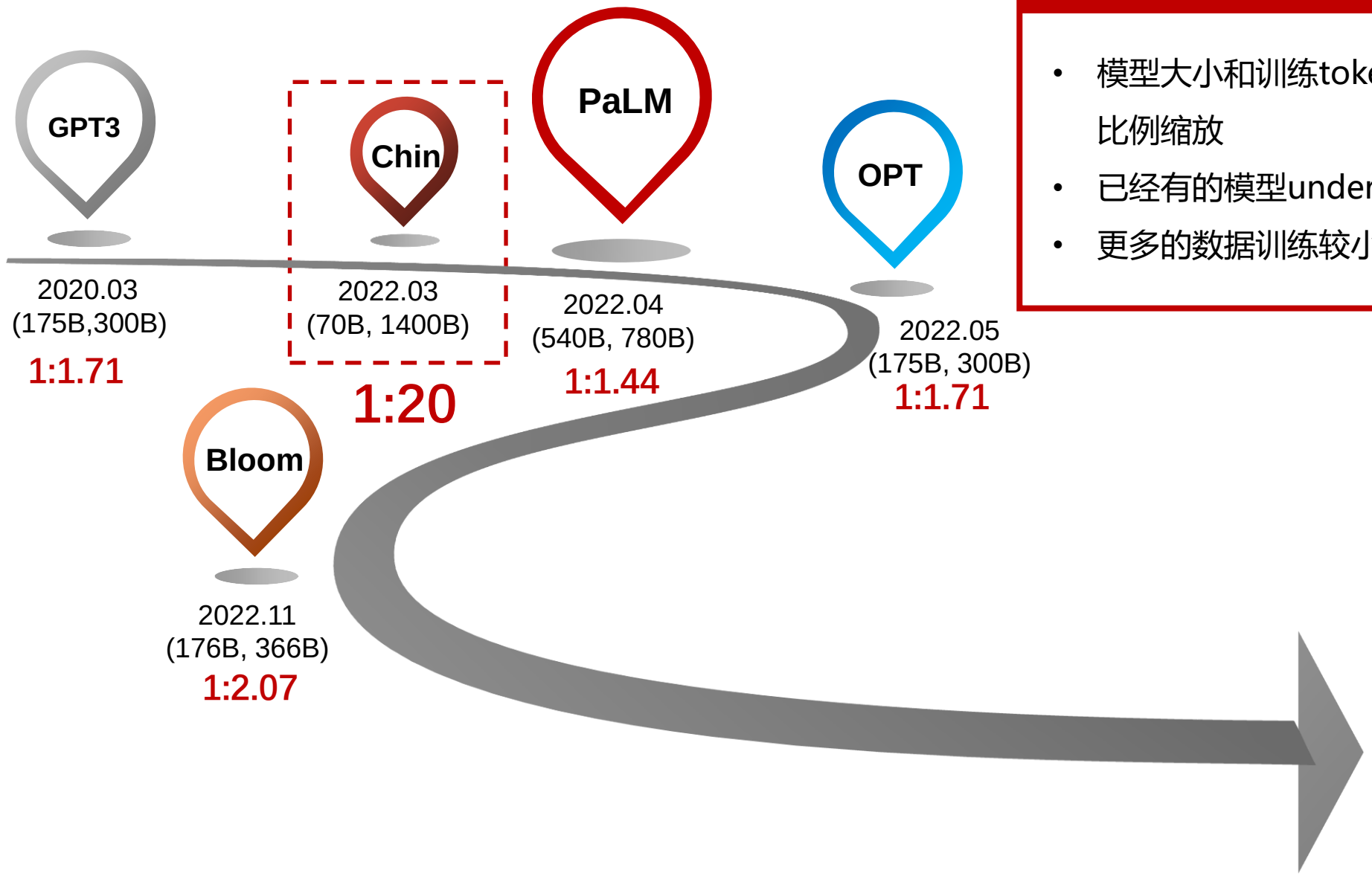
(1)你认为没有其他人会解决它。(2)你在这个主题上有独特的贡献能力。”

案例分享： “Llama的成功”



OPT: Open Pre-trained Transformer Language Models (Zhang et. Al)

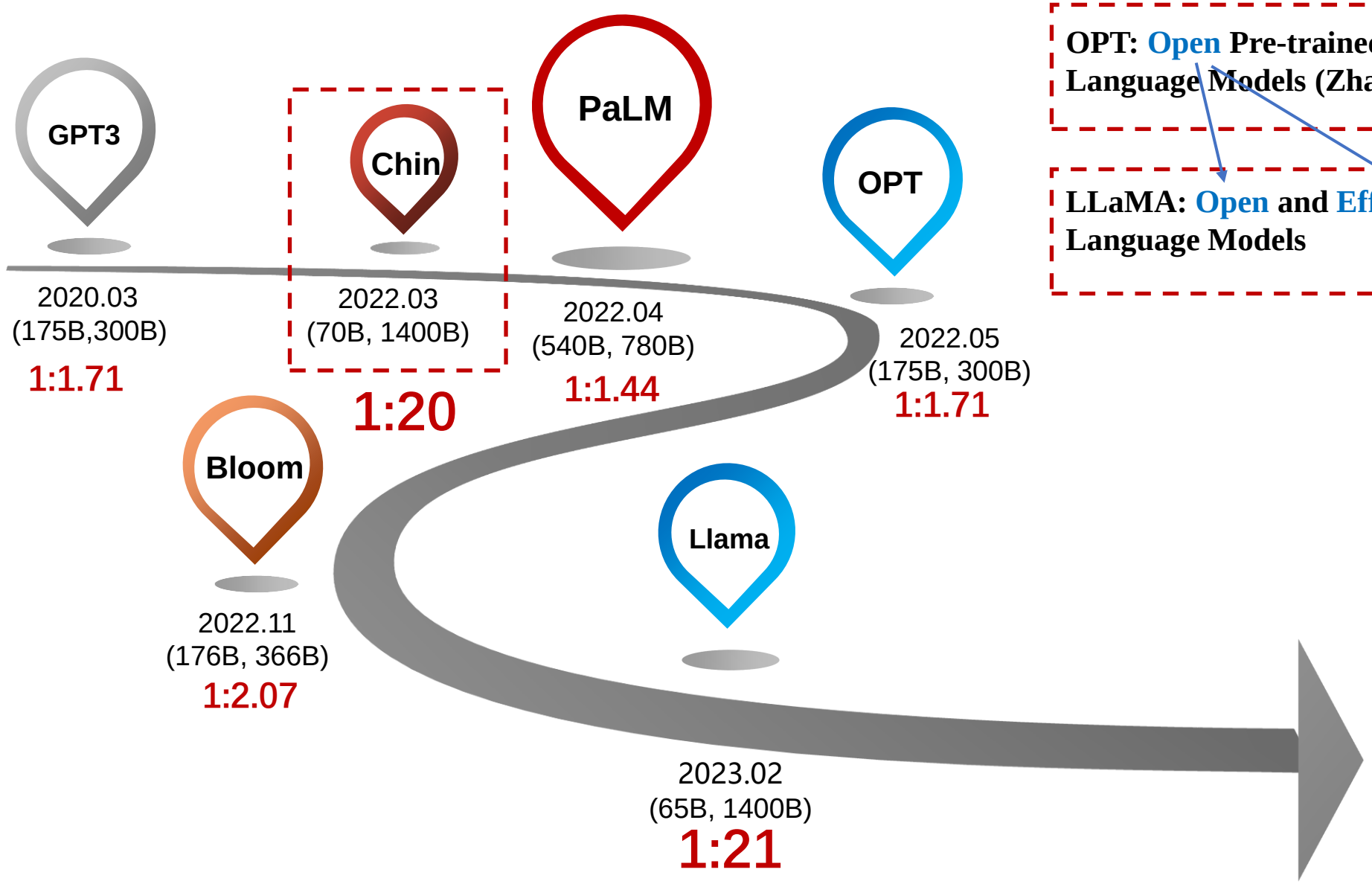
案例分享：“Llama的成功”



Chinchilla Scaling Law

- 模型大小和训练token的数量应该按相等比例缩放
- 已经有的模型under-trained (over-sized)
- 更多的数据训练较小的模型表现更好

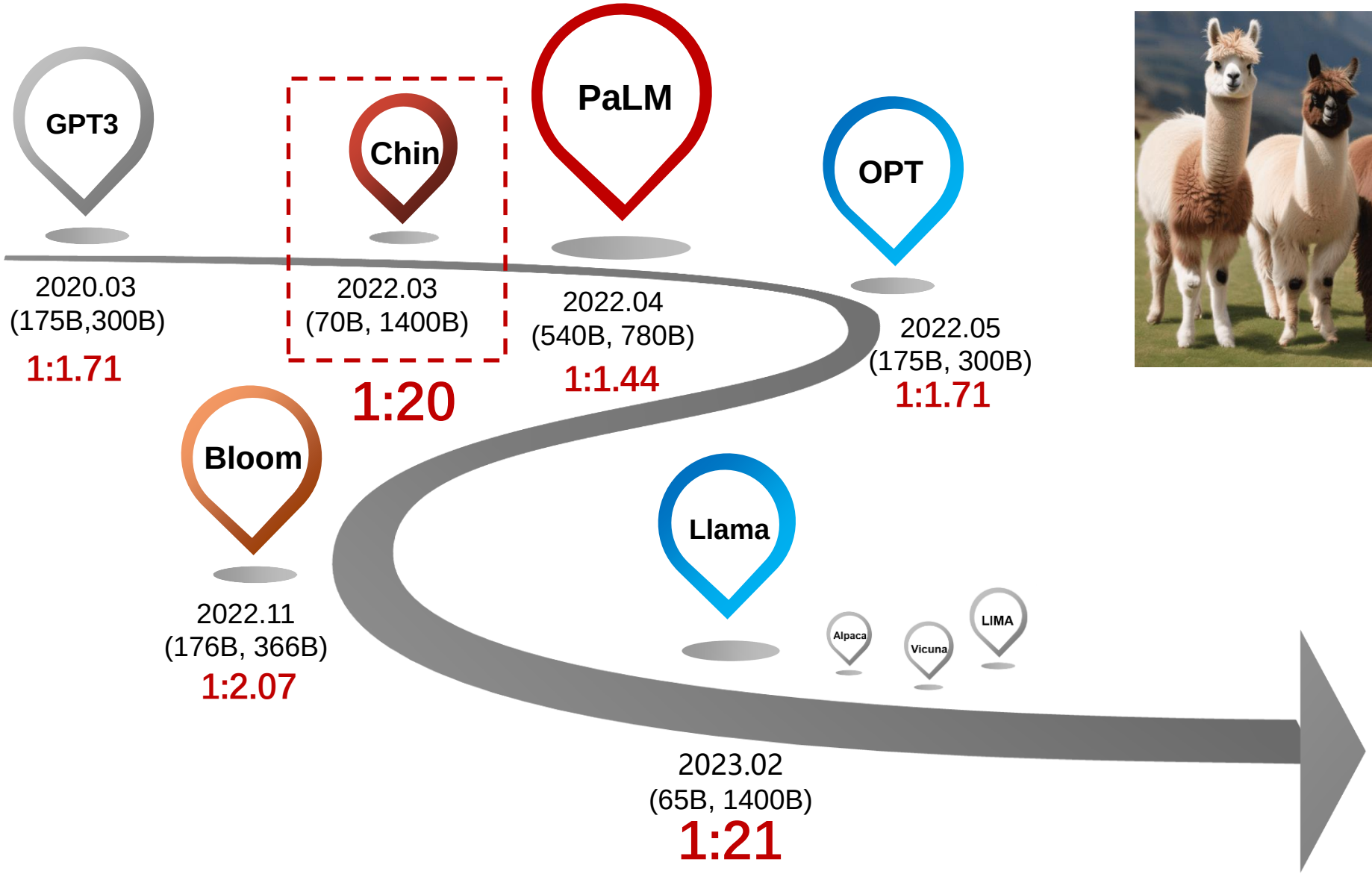
案例分享： “Llama的成功”



OPT: Open Pre-trained Transformer Language Models (Zhang et. Al)

LLaMA: Open and Efficient Foundation Language Models

案例分享： “Llama的成功”



案例分享： “Llama的成功”