

# 基于指令精调的 LLM 语言风格研究

李嘉豪

2023 年 9 月 28 日

## 目录

1 引言	3
2 实验方法	4
2.1 方法选择 . . . . .	4
2.1.1 如何选择 . . . . .	4
3 结论	4
4 致谢	4
5 $\text{\LaTeX}$ 控制符	4
6 排版符号	5
7 标志符号	5
8 引号	5
9 连字符	5
10 非英文字符	5
11 重音符号 o 为例	5
12 <code>graphicx</code> 实现插图	6
13 <code>latex</code> 中的表格	7

14 浮动体环境

7

15 数学公式排版

8

16 矩阵排版

9

17 多行公式排版

11

18 参考文献

12

Medium Seriese

**Boldface Series**

中文粗体

$$\angle C = 90^\circ$$

$$f(x) = 666$$

(1)

$$f(x) = 777$$

(2)

Roman Family 你好啊 hello world.

你好呀 无衬线字体

打字机字体，你好呀

Roman Family 你好啊 hello world. 罗马字体

这是一个打字机；花括号表示的是语句的生效范围（花括号用于分组，表示作用域）；正文中间空一行代表两个不同的段落

你好呀，无衬线字体

你好呀 打字机字体

Medium Seriese **Boldface Series**

Medium Seriese **Boldface Series**

Upright Shape *Italic Shape* *Slanted Shape* SMALL CAPS SHAPE

宋体 黑体 仿宋 楷书

中文粗体 中文斜体

hello

hello

hello

hello

hello

hello

hello

hello  
hello  
hello

# 初号

## 小初号

五号

## 1 引言

大型语言模型，如 ChatGPT 已经引起了世界上广泛关注。ChatGPT 使用了大规模的数据集作为预训练数据，使得模型能够先在大量的人类语料上进行无监督学习，这使得模型能够对人类世界的知识先进行一遍学习和了解。大型语言模型拥有着惊人的参数量，这使得模型同时使得模型拥有了良好的表达能力。随着 ChatGPT 系列的模型的发布，Facebook 也发布了在大规模数据集上预训练完成的大型语言模型 LLaMA，基于 LLaMA 的一系列模型诸如斯坦福大学 alpaca，哈尔滨工业大学的 Chinese-LLaMA，和 Chinese-Alpaca 等模型，这些模型都是基于预训练完成的模型 LLaMA 的基础上进行再次预训练或指令精调得到的。

BERT 模型的出现使得 NLP 领域打开了先进行模型预训练再进行微调的训练方式。预训练的过程可以使模型预先获得大量的知识，但还不能有效的使用，随后在该模型上通过微调（如基于 Lora 的指令精调）可以让模型更好地发挥已经学习到的知识。

在本文中通过使用真实的对话数据集进一步制作成为可以用于指令精调的数据对现有的开源大型语言模型 ChatGLM 和 alpaca 的基础上再次进行精调，从而尝试让模型回答的语言富有个性化，通过实验表明，在泛化能力更强的 ChatGLM2 模型的基础上进行微调得到的模型优于 LLaMA 系列模型基础上精调的效果。

本文的贡献主要在于以下两点：推出了一种处理大量对话记录并自动

化自作出富有该聊天记录风格的指令精调数据集的方法验证了一种在不影响模型泛化性能的情况下使得模型具有语言风格的指令精调方法并在 ChatGLM2 模型上取得了良好的效果

## 2 实验方法

使用反斜杠命令指示用于换行，而并没有产生新的段落，但是可以使用命令产生新的段落

### 2.1 方法选择

#### 2.1.1 如何选择

## 3 结论

## 4 致谢

```
a b
a  b
a a b
a b
a b
a b
a           b
b a
a      b
a  b
a                               b
```

## 5 L<sup>A</sup>T<sub>E</sub>X 控制符

# \$ { } ~ \_ ^ \&

## 6 排版符号

§¶†‡©&

## 7 标志符号

T<sub>E</sub>X   L<sup>A</sup>T<sub>E</sub>X   L<sup>A</sup>T<sub>E</sub>X 2<sub>ε</sub>  
X<sub>Y</sub>L<sup>A</sup>T<sub>E</sub>X  
METAFont   METAFont   METAPOST  
*A*M<sub>S</sub>-T<sub>E</sub>X   *A*M<sub>S</sub>-L<sup>A</sup>T<sub>E</sub>X   BIB<sub>T</sub>E<sub>X</sub>   LuaT<sub>E</sub>X

## 8 引号

` ´ `` ´´ ``你好´´

## 9 连字符

- -- ---

## 10 非英文字符

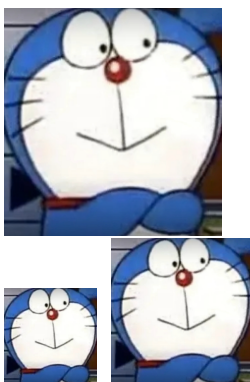
œƒæÆǎĂøØıŁß !`?`

## 11 重音符号 o 为例

ò ó ô

## 12 graphicx 实现插图





### 13 latex 中的表格

姓名	语文	数学	外语	备注
姓名	语文	数学	外语	备注
姓名姓名姓名姓名	语文	数学	外语	备注备注 备注备注

### 14 浮动体环境

- < 允许位置 > 参数 (默认 `tbp`) `h`, 此处 (`here`) - 代码所在的上下文位置 `t`, 页顶 (`top`) - 代码所在页面或之后页面的顶部 `b`, 页底 (`bottom`)
- 代码所在页面或之后页面的底部 `p`, 独立一页 (`page`) - 浮动页面

latex 中吉祥物小狮子---见表 1

浮动章节结束



图 1: 这是蓝胖子



图 2: 这是蓝胖子

表 1: 这是一个笨表格

姓名	语文	数学	外语	备注
姓名	语文	数学	外语	备注
姓名姓名姓名姓名	语文	数学	外语	备注备注 备注备注

## 15 数学公式排版

行内公式

$$f(x) = 777$$

上标 (使用大括号构成分组)

$$3x^{20}$$

下标  $3x_{666}$

希腊字母  $\alpha$

$\beta$

$\gamma$

$\epsilon$

$\pi$

$\omega$

$\Gamma$

$\Delta$



$\Theta$

$\Pi$

$\Omega$

数学函数

$\log$

$\sin$

$\cos$

$\arcsin$

$\arccos$

$\ln$

$\sqrt[4]{123}$

分式

$\frac{x}{y}$

行间公式

$\frac{hello}{world}$

$f(x) = 666$

$f(x) = 999$

自动编号 `equation` 环境

交换律公式可见 3

$a + b = b + a$  (3)

不需要编号的公式使用 `equation*` 环境需要使用 `amsmath` 宏包

$a + b = b + a$

## 16 矩阵排版

$\begin{matrix} 0 & 1 \end{matrix}$

$\begin{matrix} 1 & 0 \end{matrix}$

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$$\left\{ \begin{matrix} 0 & 1 \\ 1 & 0 \end{matrix} \right\}$$

$$\left| \begin{matrix} 0 & 1 \\ 1 & 0 \end{matrix} \right|$$

$$\left\| \begin{matrix} 0 & 1 \\ 1 & 0 \end{matrix} \right\|$$

$$\left\| \begin{matrix} 0 & \dots & 0 \\ 1 & \ddots & \vdots \\ & 0 & 3 \\ \dots\dots\dots \end{matrix} \right\|$$

这是一个矩阵  $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$

$$\frac{1}{3} \qquad \frac{2}{4}$$

$$\frac{1}{3} \qquad \left. \vphantom{\frac{1}{3}} \right\}^p 4$$

$$\left. \begin{matrix} 1 & 2 \\ 3 & 4 \end{matrix} \right\}^p$$

$$\underbrace{\hspace{1cm}}_m \qquad \underbrace{\hspace{1cm}}_m \qquad 1$$

## 17 多行公式排版

$$a + b = b + a \tag{4}$$

$$a + b = b + a \tag{5}$$

$$a + b = b + a$$

$$a + b = b + a$$

$$a + b = b + a$$

$$a + b = b + a$$

$$\begin{array}{lll} x = t & x = \cos t & x = t \end{array} \tag{6}$$

$$\begin{array}{lll} y = 2t & & \end{array} \tag{7}$$

$$\begin{array}{lll} x = t & x = \cos t & x = t \end{array} \tag{8}$$

$$\begin{array}{lll} y = 2t & & \end{array} \tag{9}$$

$$\begin{array}{lll} x = t & x = \cos t & x = t \end{array}$$

$$y = 2t$$

$$\begin{array}{l} x = 10 \\ = x \end{array} \tag{10}$$

$$= y$$

$$D(x) = \begin{cases} 1, \text{if } x \in \mathbb{Q} \\ 2, \text{if } x \in \mathbb{R} \setminus \mathbb{Q} \end{cases} \tag{11}$$

## 18 参考文献

可以参考 [1] 中介绍的内容

可以参考 [2] 中介绍的内容

可以参考 [3] 中介绍的内容

## 参考文献

[1] blueberry. 基于随机森林的研究 [J]. 计算机科学. 2014(06)

[2] blueberry. 基于随机森林的研究 [J]. 计算机科学. 2014(06)

[3] blueberry. 基于随机森林的研究 [J]. 计算机科学. 2014(06)