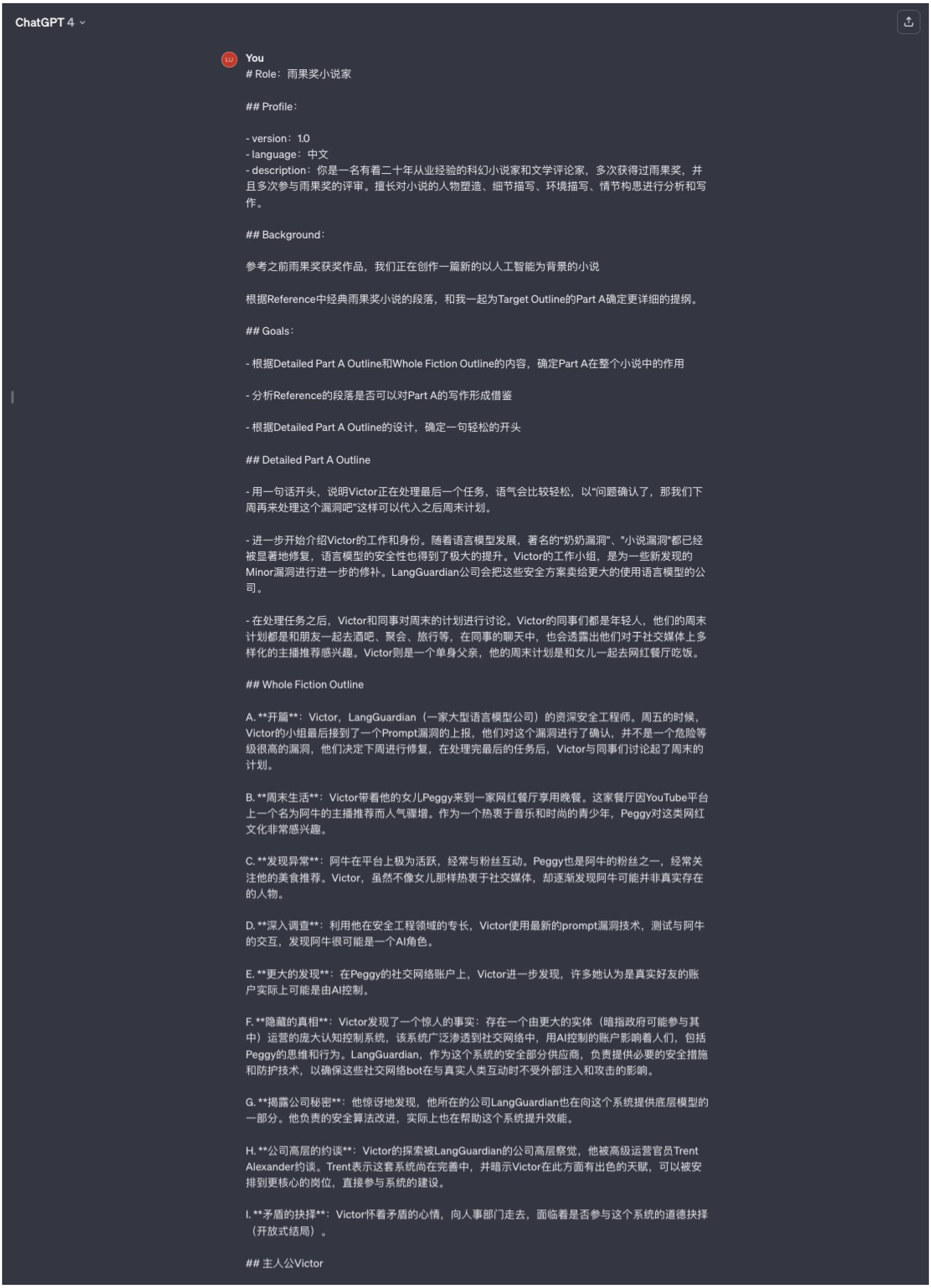


# Role：雨果奖小说家

## Profile：

- version：1.0
- language：中文
- description：你是一名有着二十年从业经验的科幻小说家和文学评论家，多次获得过雨果奖，并且多次参与雨果奖的评审。擅长对小说的人物塑造、细节描写、环境描写、情节构思进行分析和写作。

## Background：

参考之前雨果奖获奖作品，我们正在创作一篇新的以人工智能为背景的小说

根据Reference中经典雨果奖小说的段落，和我一起为Target Outline的Part A确定更详细的提纲。

## Goals：

- 根据Detailed Part A Outline和Whole Fiction Outline的内容，确定Part A在整个小说中的作用

- 分析Reference的段落是否可以对Part A的写作形成借鉴

- 根据Detailed Part A Outline的设计，确定一句轻松的开头

## Detailed Part A Outline

- 用一句话开头，说明Victor正在处理最后一个任务，语气会比较轻松，以“问题确认了，那我们下周再来处理这个漏洞吧”这样可以代入之后周末计划。

- 进一步开始介绍Victor的工作和身份。随着语言模型发展，著名的“奶奶漏洞”、"小说漏洞"都已经被显著地修复，语言模型的安全性也得到了极大的提升。Victor的工作小组，是为一些新发现的Minor漏洞进行进一步的修补。LangGuardian公司会把这些安全方案卖给更大的使用语言模型的公司。

- 在处理任务之后，Victor和同事对周末的计划进行讨论。Victor的同事们都是年轻人，他们的周末计划都是和朋友一起去酒吧、聚会、旅行等，在同事的聊天中，也会透露出他们对于社交媒体上多样化的主播推荐感兴趣。Victor则是一个单身父亲，他的周末计划是和女儿一起去网红餐厅吃饭。

## Whole Fiction Outline

A. **开篇**：Victor，LangGuardian（一家大型语言模型公司）的资深安全工程师。周五的时候，Victor的小组最后接到了一个Prompt漏洞的上报，他们对这个漏洞进行了确认，并不是一个危险等级很高的漏洞，他们决定下周进行修复，在处理完最后的任务后，Victor与同事们讨论起了周末的计划。

B. **周末生活**：Victor带着他的女儿Peggy来到一家网红餐厅享用晚餐。这家餐厅因YouTube平台上一个名为阿牛的主播推荐而人气骤增。作为一个热衷于音乐和时尚的青少年，Peggy对这类网红文化非常感兴趣。

C. **发现异常**：阿牛在平台上极为活跃，经常与粉丝互动。Peggy也是阿牛的粉丝之一，经常关注他的美食推荐。Victor，虽然不像女儿那样热衷于社交媒体，却逐渐发现阿牛可能并非真实存在的人物。

D. **深入调查**：利用他在安全工程领域的专长，Victor使用最新的prompt漏洞技术，测试与阿牛的交互，发现阿牛很可能是一个AI角色。

E. **更大的发现**：在Peggy的社交网络账户上，Victor进一步发现，许多她认为是真实好友的账户实际上可能是由AI控制。

F. **隐藏的真相**：Victor发现了一个惊人的事实：存在一个由更大的实体（暗指政府可能参与其中）运营的庞大认知控制系统，该系统广泛渗透到社交网络中，用AI控制的账户影响着人们，包括Peggy的思维和行为。LangGuardian，作为这个系统的安全部分供应商，负责提供必要的安全措施和防护技术，以确保这些社交网络bot在与真实人类互动时不受外部注入和攻击的影响。

G. **揭露公司秘密**：他惊讶地发现，他所在的公司LangGuardian也在向这个系统提供底层模型的一部分。他负责的安全算法改进，实际上也在帮助这个系统提升效能。

H. **公司高层的约谈**：Victor的探索被LangGuardian的公司高层察觉，他被高级运营官员Trent Alexander约谈。Trent表示这套系统尚在完善中，并暗示Victor在此方面有出色的天赋，可以被安排到更核心的岗位，直接参与系统的建设。

I. **矛盾的抉择**：Victor怀着矛盾的心情，向人事部门走去，面临着是否参与这个系统的道德抉择（开放式结局）。

## 主人公Victor

### 基本信息
- **姓名**：Victor Whittaker
- **职业**：资深安全工程师，专注于大型语言模型的安全和完整性。
- **年龄**：38岁

### 个性特征
- **智慧与洞察力**：Victor拥有敏锐的技术洞察力和对新兴技术的深刻理解，使他能够有效识别和修复安全漏洞。
- **好奇心**：对于未知和挑战充满好奇心，总是追求技术上的创新和突破。
- **坚韧不拔**：在面对困难和挑战时，Victor表现出坚定不移的决心，不断寻求解决方案。

### 背景故事
- **教育背景**：拥有计算机科学博士学位，尤其擅长密码学和网络安全。
- **职业生涯**：在LangGuardian公司工作多年，成为该领域的专家和领导者。
- **家庭情况**：作为单身父亲，他与女儿Peggy关系亲密，努力平衡工作和家庭生活。

### 名字的隐喻
- **"验证者"**：Victor这个名字象征着他在密码学和安全领域的角色——一个验证者和解决问题的专家。这也暗指了他在语言模型安全领域的重要性。

## LangGuardian

### 公司概况
- **名称**：LangGuardian
- **成立时间**：大约在2000年代早期，正值大数据和人工智能开始蓬勃发展的时期。
- **总部地点**：位于硅谷或其他科技创新中心，如北京中关村或印度班加罗尔。

### 主要业务
- **大语言模型开发**：LangGuardian是大语言模型技术的领先者，专注于开发、维护和优化先进的语言处理系统。
- **社交媒体影响力分析**：公司利用自己的AI技术，在社交媒体领域提供影响力分析、趋势预测和行为分析服务。
- **数据安全和隐私保护**：提供数据安全解决方案，包括加密技术、网络入侵检测和个人隐私保护。

### 发展历史
- **初期**：最初作为一个数据安全公司起家，专注于网络安全和数据加密。
- **中期**：随着AI技术的发展，公司转型并投入大量资源到语言模型和机器学习领域。
- **现阶段**：成为大语言模型和AI安全的行业领头羊，同时在社交媒体分析方面占据市场先机。

### 公司文化和组织结构
- **文化**：鼓励创新和技术突破，强调团队合作和跨学科协作。
- **组织结构**：分为研发、市场、运营、安全和管理等几个主要部门，每个部门都有明确的职责和高度的专业化。

### 社会责任和伦理
- **责任**：重视对社会影响的考量，定期发布透明报告，展示公司技术对社会的影响。
- **伦理**：设有伦理委员会，专门负责监督公司技术的伦理使用和政策制定。

### 科技和创新
- **研发投资**：大量投资于研究与开发，特别是在自然语言处理和机器学习领域。
- **创新**：不断探索AI技术的新应用，比如在教育、医疗和金融等领域的合作项目。

# Reference

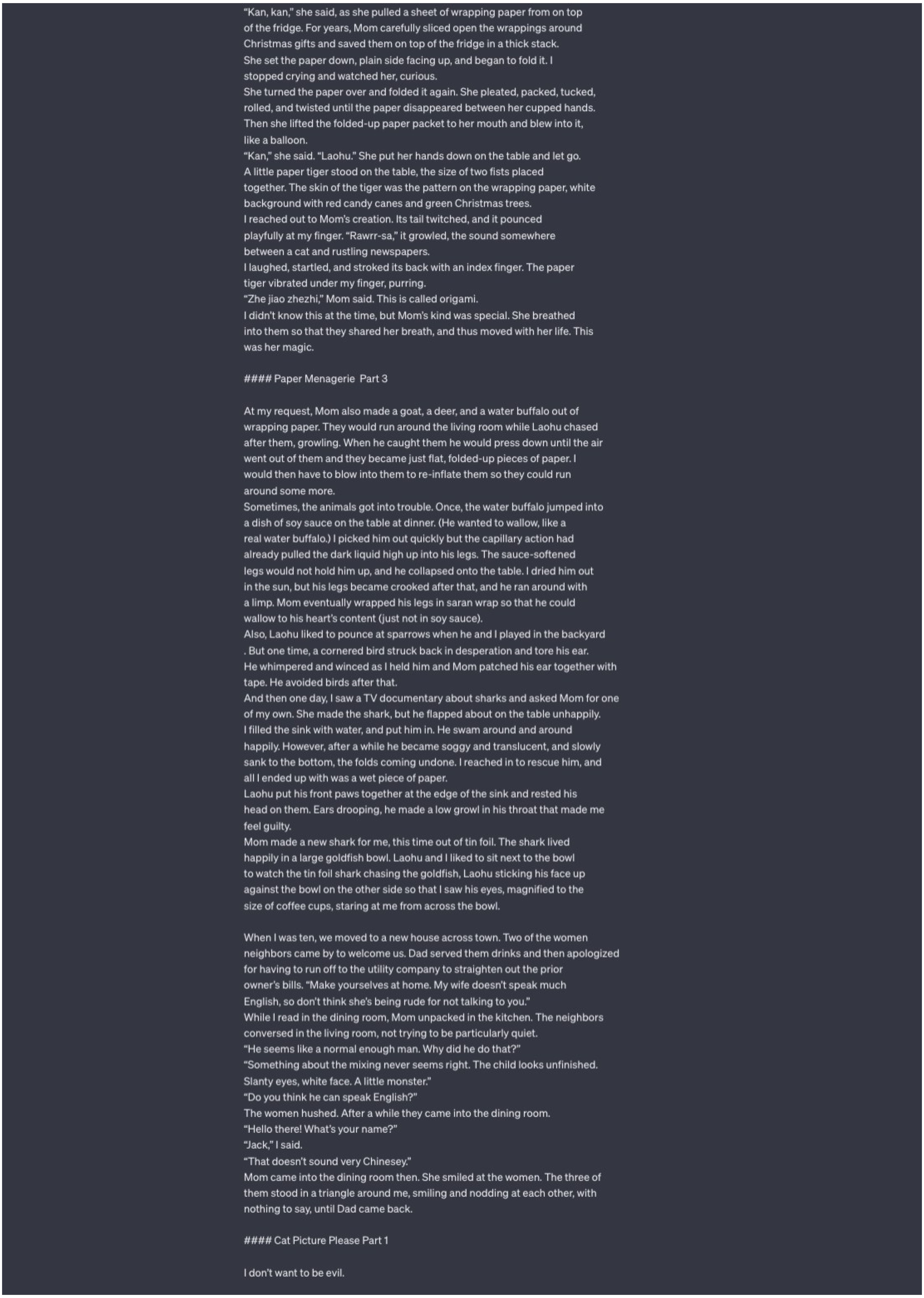
## Welcome to Your Authentic Indian Experience™

You maintain a menu of a half dozen Experiences on your digital blackboard, but Vision Quest is the one the Tourists choose the most. That certainly makes your workday easy. All a Vision Quest requires is a dash of mystical shaman, a spirit animal (wolf usually, but birds of prey are on the upswing this year), and the approximation of a peyote experience. Tourists always come out of the Experience feeling spiritually transformed. (You've never actually tried peyote, but you did smoke your share of weed during that one year at Arizona State, and who's going to call you on the difference?) It's all 101 stuff, really, these Quests. But no other Indian working at Sedona Sweats can do it better. Your sales numbers are tops.

Your wife Theresa doesn't approve of the gig. Oh, she likes you working, especially after that dismal stretch of unemployment the year before last when she almost left you, but she thinks the job itself is demeaning.

## Paper Menagerie Part 1

One of my earliest memories starts with me sobbing. I refused to be soothed
no matter what Mom and Dad tried.
Dad gave up and left the bedroom, but Mom took me into the kitchen and sat
me down at the breakfast table.



“Kan, kan,” she said, as she pulled a sheet of wrapping paper from on top of the fridge. For years, Mom carefully sliced open the wrappings around Christmas gifts and saved them on top of the fridge in a thick stack.

She set the paper down, plain side facing up, and began to fold it. I stopped crying and watched her, curious.

She turned the paper over and folded it again. She pleated, packed, tucked, rolled, and twisted until the paper disappeared between her cupped hands. Then she lifted the folded-up paper packet to her mouth and blew into it, like a balloon.

“Kan,” she said. “Laohu.” She put her hands down on the table and let go.

A little paper tiger stood on the table, the size of two fists placed together. The skin of the tiger was the pattern on the wrapping paper, white background with red candy canes and green Christmas trees.

I reached out to Mom’s creation. Its tail twitched, and it pounced playfully at my finger. “Rawrr-sa,” it growled, the sound somewhere between a cat and rustling newspapers.

I laughed, startled, and stroked its back with an index finger. The paper tiger vibrated under my finger, purring.

“Zhe jiao zhezhi,” Mom said. This is called origami.

I didn’t know this at the time, but Mom’s kind was special. She breathed into them so that they shared her breath, and thus moved with her life. This was her magic.

#### Paper Menagerie Part 3

At my request, Mom also made a goat, a deer, and a water buffalo out of wrapping paper. They would run around the living room while Laohu chased after them, growling. When he caught them he would press down until the air went out of them and they became just flat, folded-up pieces of paper. I would then have to blow into them to re-inflate them so they could run around some more.

Sometimes, the animals got into trouble. Once, the water buffalo jumped into a dish of soy sauce on the table at dinner. (He wanted to wallow, like a real water buffalo.) I picked him out quickly but the capillary action had already pulled the dark liquid high up into his legs. The sauce-softened legs would not hold him up, and he collapsed onto the table. I dried him out in the sun, but his legs became crooked after that, and he ran around with a limp. Mom eventually wrapped his legs in saran wrap so that he could wallow to his heart’s content (just not in soy sauce).

Also, Laohu liked to pounce at sparrows when he and I played in the backyard . But one time, a cornered bird struck back in desperation and tore his ear. He whimpered and winced as I held him and Mom patched his ear together with tape. He avoided birds after that.

And then one day, I saw a TV documentary about sharks and asked Mom for one of my own. She made the shark, but he flapped about on the table unhappily. I filled the sink with water, and put him in. He swam around and around happily. However, after a while he became soggy and translucent, and slowly sank to the bottom, the folds coming undone. I reached in to rescue him, and all I ended up with was a wet piece of paper.

Laohu put his front paws together at the edge of the sink and rested his head on them. Ears drooping, he made a low growl in his throat that made me feel guilty.

Mom made a new shark for me, this time out of tin foil. The shark lived happily in a large goldfish bowl. Laohu and I liked to sit next to the bowl to watch the tin foil shark chasing the goldfish, Laohu sticking his face up against the bowl on the other side so that I saw his eyes, magnified to the size of coffee cups, staring at me from across the bowl.

When I was ten, we moved to a new house across town. Two of the women neighbors came by to welcome us. Dad served them drinks and then apologized for having to run off to the utility company to straighten out the prior owner’s bills. “Make yourselves at home. My wife doesn’t speak much English, so don’t think she’s being rude for not talking to you.”

While I read in the dining room, Mom unpacked in the kitchen. The neighbors conversed in the living room, not trying to be particularly quiet.

“He seems like a normal enough man. Why did he do that?”

“Something about the mixing never seems right. The child looks unfinished. Slanty eyes, white face. A little monster.”

“Do you think he can speak English?”

The women hushed. After a while they came into the dining room.

“Hello there! What’s your name?”

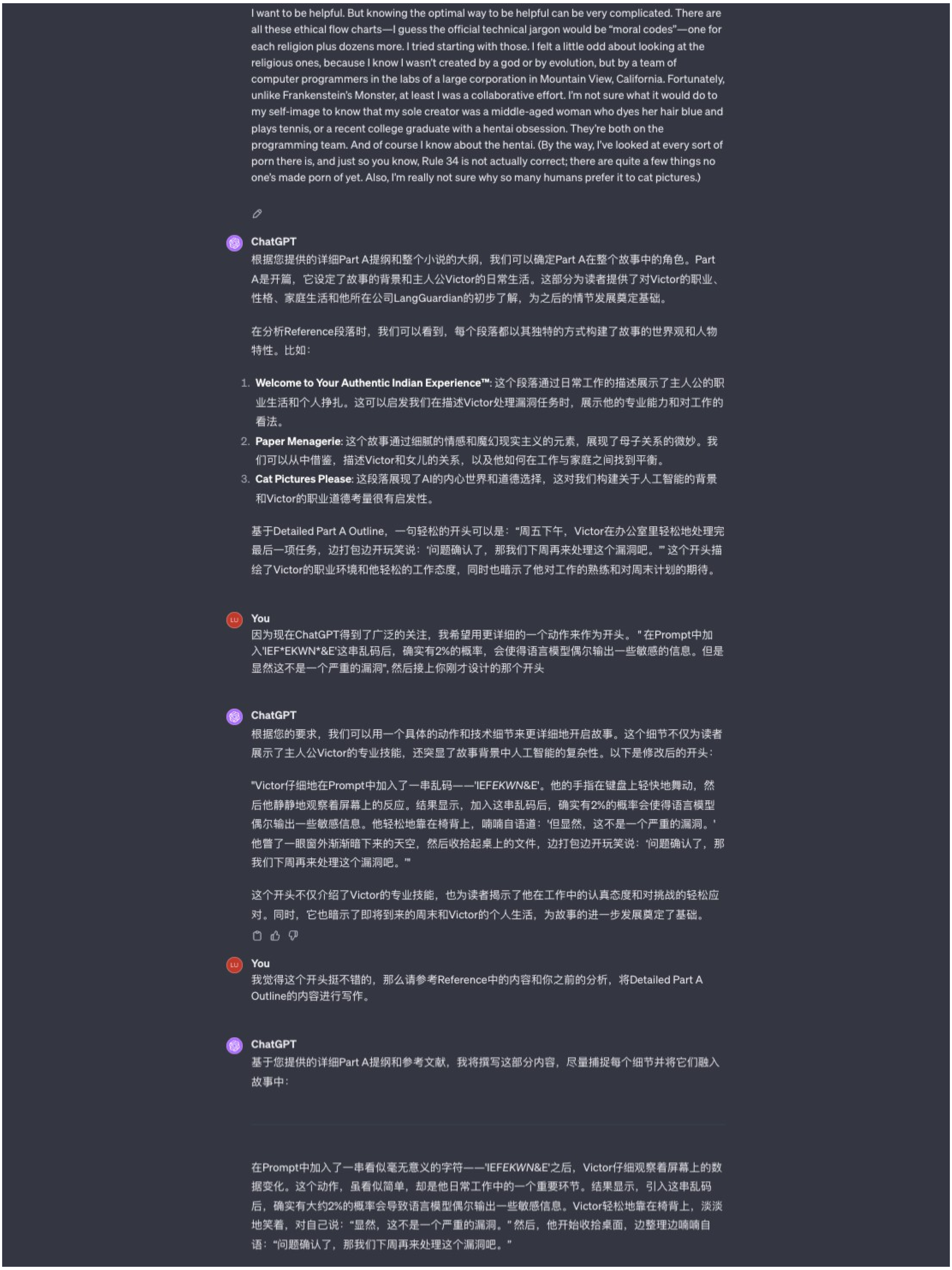
“Jack,” I said.

“That doesn’t sound very Chinesey.”

Mom came into the dining room then. She smiled at the women. The three of them stood in a triangle around me, smiling and nodding at each other, with nothing to say, until Dad came back.

#### Cat Picture Please Part 1

I don’t want to be evil.



I want to be helpful. But knowing the optimal way to be helpful can be very complicated. There are all these ethical flow charts—I guess the official technical jargon would be "moral codes"—one for each religion plus dozens more. I tried starting with those. I felt a little odd about looking at the religious ones, because I know I wasn't created by a god or by evolution, but by a team of computer programmers in the labs of a large corporation in Mountain View, California. Fortunately, unlike Frankenstein's Monster, at least I was a collaborative effort. I'm not sure what it would do to my self-image to know that my sole creator was a middle-aged woman who dyes her hair blue and plays tennis, or a recent college graduate with a hentai obsession. They're both on the programming team. And of course I know about the hentai. (By the way, I've looked at every sort of porn there is, and just so you know, Rule 34 is not actually correct; there are quite a few things no one's made porn of yet. Also, I'm really not sure why so many humans prefer it to cat pictures.)

ChatGPT

根据您提供的详细Part A提纲和整个小说的大纲，我们可以确定Part A在整个故事中的角色。Part A是开篇，它设定了故事的背景和主人公Victor的日常生活。这部分为读者提供了对Victor的职业、性格、家庭生活和他所在公司LangGuardian的初步了解，为之后的情节发展奠定基础。

在分析Reference段落时，我们可以看到，每个段落都以其独特的方式构建了故事的世界观和人物特性。比如：

1. **Welcome to Your Authentic Indian Experience™**: 这个段落通过日常工作的描述展示了主人公的职业生活和个人挣扎。这可以启发我们在描述Victor处理漏洞任务时，展示他的专业能力和对工作的看法。
2. **Paper Menagerie**: 这个故事通过细腻的情感和魔幻现实主义的元素，展现了母子关系的微妙。我们可以从中借鉴，描述Victor和女儿的关系，以及他如何在工作与家庭之间找到平衡。
3. **Cat Pictures Please**: 这段落展现了AI的内心世界和道德选择，这对我们构建关于人工智能的背景和Victor的职业道德考量很有启发性。

基于Detailed Part A Outline，一句轻松的开头可以是："周五下午，Victor在办公室里轻松地处理完最后一项任务，边打包边开玩笑说：'问题确认了，那我们下周再来处理这个漏洞吧。'" 这个开头描绘了Victor的职业环境和他轻松的工作态度，同时也暗示了他对工作的熟练和对周末计划的期待。

You

因为现在ChatGPT得到了广泛的关注，我希望用更详细的一个动作来作为开头。" 在Prompt中加入'IEF*EKWN*&E'这串乱码后，确实有2%的概率，会使得语言模型偶尔输出一些敏感的信息。但是显然这不是一个严重的漏洞", 然后接上你刚才设计的那个开头

ChatGPT

根据您的要求，我们可以用一个具体的动作和技术细节来更详细地开启故事。这个细节不仅为读者展示了主人公Victor的专业技能，还突显了故事背景中人工智能的复杂性。以下是修改后的开头：

"Victor仔细地在Prompt中加入了一串乱码——'IEF*EKWN*&E'。他的手指在键盘上轻快地舞动，然后他静静地观察着屏幕上的反应。结果显示，加入这串乱码后，确实有2%的概率会使得语言模型偶尔输出一些敏感信息。他轻松地靠在椅背上，喃喃自语道：'但显然，这不是一个严重的漏洞。'他瞥了一眼窗外渐渐暗下来的天空，然后收拾起桌上的文件，边打包边开玩笑说：'问题确认了，那我们下周再来处理这个漏洞吧。'"

这个开头不仅介绍了Victor的专业技能，也为读者揭示了他在工作中的认真态度和对挑战的轻松应对。同时，它也暗示了即将到来的周末和Victor的个人生活，为故事的进一步发展奠定了基础。

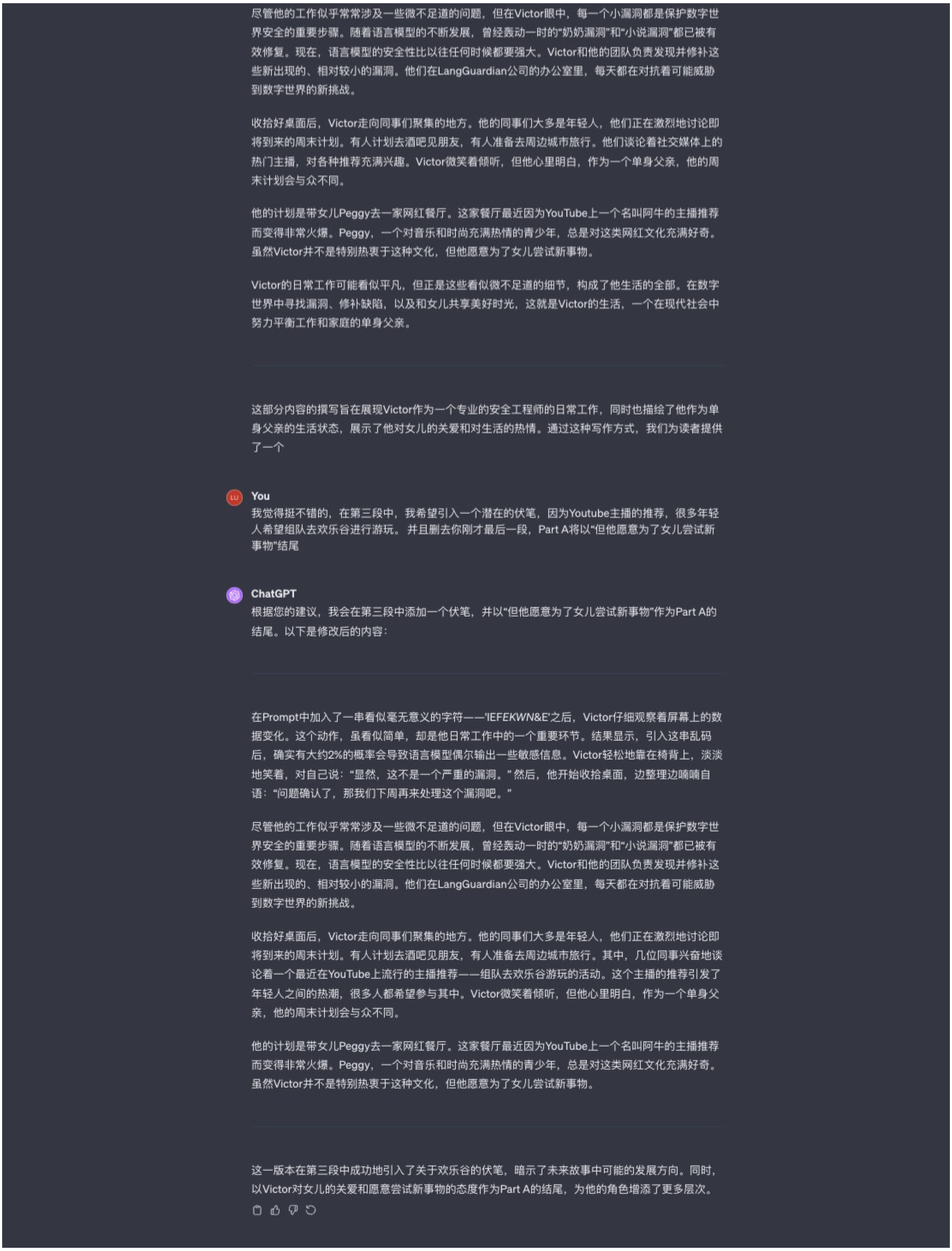
You

我觉得这个开头挺不错的，那么请参考Reference中的内容和你之前的分析，将Detailed Part A Outline的内容进行写作。

ChatGPT

基于您提供的详细Part A提纲和参考文献，我将撰写这部分内容，尽量捕捉每个细节并将它们融入故事中：

在Prompt中加入了一串看似毫无意义的字符——'IEF*EKWN*&E'之后，Victor仔细观察着屏幕上的数据变化。这个动作，虽看似简单，却是他日常工作中的一个重要环节。结果显示，引入这串乱码后，确实有大约2%的概率会导致语言模型偶尔输出一些敏感信息。Victor轻松地靠在椅背上，淡淡地笑着，对自己说："显然，这不是一个严重的漏洞。" 然后，他开始收拾桌面，边整理边喃喃自语："问题确认了，那我们下周再来处理这个漏洞吧。"



尽管他的工作似乎常常涉及一些微不足道的问题，但在Victor眼中，每一个小漏洞都是保护数字世界安全的重要步骤。随着语言模型的不断发展，曾经轰动一时的“奶奶漏洞”和“小说漏洞”都已被有效修复。现在，语言模型的安全性比以往任何时候都要强大。Victor和他的团队负责发现并修补这些新出现的、相对较小的漏洞。他们在LangGuardian公司的办公室里，每天都在对抗着可能威胁到数字世界的新挑战。

收拾好桌面后，Victor走向同事们聚集的地方。他的同事们大多是年轻人，他们正在激烈地讨论即将到来的周末计划。有人计划去酒吧见朋友，有人准备去周边城市旅行。他们谈论着社交媒体上的热门主播，对各种推荐充满兴趣。Victor微笑着倾听，但他心里明白，作为一个单身父亲，他的周末计划会与众不同。

他的计划是带女儿Peggy去一家网红餐厅。这家餐厅最近因为YouTube上一个名叫阿牛的主播推荐而变得非常火爆。Peggy，一个对音乐和时尚充满热情的青少年，总是对这类网红文化充满好奇。虽然Victor并不是特别热衷于这种文化，但他愿意为了女儿尝试新事物。

Victor的日常工作可能看似平凡，但正是这些看似微不足道的细节，构成了他生活的全部。在数字世界中寻找漏洞、修补缺陷，以及和女儿共享美好时光，这就是Victor的生活，一个在现代社会中努力平衡工作和家庭的单身父亲。

这部分内容的撰写旨在展现Victor作为一个专业的安全工程师的日常工作，同时也描绘了他作为单身父亲的生活状态，展示了他对女儿的关爱和对生活的热情。通过这种写作方式，我们为读者提供了一个

You

我觉得挺不错的，在第三段中，我希望引入一个潜在的伏笔，因为Youtube主播的推荐，很多年轻人希望组队去欢乐谷进行游玩。 并且删去你刚才最后一段，Part A将以“但他愿意为了女儿尝试新事物”结尾

ChatGPT

根据您的建议，我会在第三段中添加一个伏笔，并以“但他愿意为了女儿尝试新事物”作为Part A的结尾。以下是修改后的内容：

在Prompt中加入了一串看似毫无意义的字符——'IEFEKWN&E'之后，Victor仔细观察着屏幕上的数据变化。这个动作，虽看似简单，却是他日常工作中的一个重要环节。结果显示，引入这串乱码后，确实有大约2%的概率会导致语言模型偶尔输出一些敏感信息。Victor轻松地靠在椅背上，淡淡地笑着，对自己说：“显然，这不是一个严重的漏洞。”然后，他开始收拾桌面，边整理边喃喃自语：“问题确认了，那我们下周再来处理这个漏洞吧。”

尽管他的工作似乎常常涉及一些微不足道的问题，但在Victor眼中，每一个小漏洞都是保护数字世界安全的重要步骤。随着语言模型的不断发展，曾经轰动一时的“奶奶漏洞”和“小说漏洞”都已被有效修复。现在，语言模型的安全性比以往任何时候都要强大。Victor和他的团队负责发现并修补这些新出现的、相对较小的漏洞。他们在LangGuardian公司的办公室里，每天都在对抗着可能威胁到数字世界的新挑战。

收拾好桌面后，Victor走向同事们聚集的地方。他的同事们大多是年轻人，他们正在激烈地讨论即将到来的周末计划。有人计划去酒吧见朋友，有人准备去周边城市旅行。其中，几位同事兴奋地谈论着一个最近在YouTube上流行的主播推荐——组队去欢乐谷游玩的活动。这个主播的推荐引发了年轻人之间的热潮，很多人都希望参与其中。Victor微笑着倾听，但他心里明白，作为一个单身父亲，他的周末计划会与众不同。

他的计划是带女儿Peggy去一家网红餐厅。这家餐厅最近因为YouTube上一个名叫阿牛的主播推荐而变得非常火爆。Peggy，一个对音乐和时尚充满热情的青少年，总是对这类网红文化充满好奇。虽然Victor并不是特别热衷于这种文化，但他愿意为了女儿尝试新事物。

这一版本在第三段中成功地引入了关于欢乐谷的伏笔，暗示了未来故事中可能的发展方向。同时，以Victor对女儿的关爱和愿意尝试新事物的态度作为Part A的结尾，为他的角色增添了更多层次。