

# 北京郵電大學

## 本科 毕业 设计 (论文)



题目： 社猜猜看这个毕设题目是什么

姓 名 猜 猜

学 院 信息与通信工程学院

专 业 通信工程

班 级 201421199

学 号 2014210999

班内序号 99

指导教师 猜 猜

2018 年 5 月

# 北京邮电大学

## 本科毕业设计（论文）任务书

学院	信息与通信工程学院	专业	通信工程	班级	2014211199
学生姓名	猜猜	学号	2014210999	班内序号	99
指导教师姓名	猜猜	所在单位	信息与通信工程学院	职称	教授
设计(论文)题目	(中文) 猜猜看毕设题目是什么				
	(英文) Just Guess What On Earth My Title is				
题目分类	工程实践类 <input type="checkbox"/> 研究设计类 <input checked="" type="checkbox"/> 理论分析类 <input type="checkbox"/>				
题目来源	题目是否来源于科研项目      是 <input type="checkbox"/> 否 <input checked="" type="checkbox"/>				
	科研项目名称:				
	科研项目负责人:				
<p>主要任务及目标:</p> <ul style="list-style-type: none"><li>给出一个 LaTeX 模板。</li><li>减小大家花费在排版上的时间成本。</li><li>但是代价是需要有一点基础学习成本，但研究生时可能会用到。</li></ul>					
<p>主要内容:</p> <p>同任务目标。</p>					
<p>主要参考文献:</p> <ul style="list-style-type: none"><li>Zubiaga A, Aker A, Bontcheva K, et al. Detection and Resolution of Rumours in Social Media: A Survey[J]. 2017.</li><li>Yan Z, Chen W, Chai K Y, et al. Detecting rumors on Online Social Networks using multi-layer autoencoder[C]// IEEE Technology &amp; Engineering Management Conference. IEEE, 2017:437-441.</li></ul>					
<p>进度安排:</p> <ul style="list-style-type: none"><li>2018.1.1 ~ 2018.2.10 完成领域内容调研，模板对应部分撰写。</li><li>2018.2.28~2018.4.15 完成相关模板研究，设计模板。</li><li>2018.4.16~2018.4.30 进行模板设计评估和比较分析。</li><li>2018.5.1~2018.5.15 模板整体撰写。</li></ul>					
指导教师签字		日期	年 月 日		

编号: \_\_\_\_\_

## 北京邮电大学本科生毕业设计（论文）成绩评定表

学生姓名	猜 猜		所在学院	信息与通信工程学院					
学号	2014210999	专业	通信工程	班级	2014211199				
论文题目	(中文) 猜猜看毕设题目是什么								
	(英文) Just Guess What On Earth My Title Is								
指导教师姓名		指导教师职称		指导教师单位					
中期检查小组评分	(满分 10 分):			中期检查小组组长签字:		检查日期:			
指导教师评分	评价内容	具体要求			分值			评分	
	调研论证	能独立查阅文献和从事相关调研；能正确翻译外文资料；有收集、加工各种信息及获取新知识的能力和自学能力。			5	4	3.5	3	2
	方案设计	能独立提出符合选题的可行性研究方案、实验方案、设计方案，独立进行实验（如安装、调试、操作）和研究方案论证。			5	4	3.5	3	2
	能力水平	能综合运用所学知识和技能去分析与解决毕业设计（论文）过程中遇到的实际问题；能正确处理实验数据；能对课题进行理论分析，得出有价值的结论。			5	4	3.5	3	2
	学习态度	认真、勤奋、努力、诚实、严格遵守纪律，按期饱满完成规定的任务。			5	4	3.5	3	2
	设计（论文）水平	文题相符、综述简练完整，有见解；立论正确，论述充分，结论严谨合理；实验正确，分析处理科学；文字通顺，技术用语准确，设计（论文）有理论价值和应用价值。			5	4	3.5	3	2
	文本规范	装订顺序正确，字体字号等与基本规范相符，符号统一，编号齐全，图表完备、整洁、正确。			5	4	3.5	3	2
指导教师评分合计(满分 30 分): 评语:									
指导教师签字:					日期: 年 月 日				
复议	<input type="checkbox"/> 是 <input type="checkbox"/> 否    复议评分合计:                      复议人签字:                      复议日期: 复议有权限修改指导教师评分，选择复议后指导教师评分将由复议评分替换								

本科生毕业设计（论文）答辩成绩评定标准														
答辩小组成绩评定	评价内容	具体要求	分值											
	选题	符合专业培养目标，符合社会实际、结合工程实际，难易适度，体现新颖性、综合性。	5	4	3.5	3	2							
	设计（论文）质量水平	全面完成任务书中规定的各项要求，文题相符，工作量饱满，写作规范，达到综合训练的要求，有理论成果和应用价值。	20	16	14	12	8							
	答辩准备	准备充分；有简洁、清晰、美观的演示文稿；准时到场。	5	4	3.5	3	2							
	内容陈述	语言表达简洁、流利、清楚、准确，思路清晰，重点突出，逻辑性强，概念清楚，论点正确；实验方法科学，分析归纳合理；结论严谨；表现出对毕业设计（论文）内容掌握透彻。	20	18	14	12	8							
	回答问题	回答问题准确、有深度、有理论根据、基本概念清晰。	10	8	7	6	4							
	答辩小组评分合计（满分 60 分）													
意见：														
答辩小组组长签字：_____ 年 月 日														
答辩小组成员：														
学院意见	最终成绩：百分制_____； 五分制_____													
	院长签章：_____ 学院盖章：_____ 年 月 日													
备注														

注：1. 毕业设计（论文）成绩由中期检查评分（满分 10 分）、指导教师评分/复议评分（满分 30 分）和答辩小组评分（满分 60 分）相加，得出百分制成绩，再按 100-90 分为“优”、89-80 分为“良”、79-70 分为“中”、69-60 分为“及格”、60 分以下为“不及格”的标准折合成五级分制成绩；

2. 此表原件一式三份，一份存入学生档案，一份装订到毕业论文中，一份交教务处存入档案馆。

# 北京邮电大学

## 本科毕业设计（论文）诚信声明

本人声明所呈交的毕业设计（论文），题目《》是  
本人在指导教师的指导下，独立进行研究工作所取得的成果。尽我所知，除了文中特别  
加以标注和致谢中所罗列的内容以外，论文中不包含其他人已经发表或撰写过的研究成果，  
也不包含为获得北京邮电大学或其他教育机构的学位或证书而使用过的材料。

申请学位论文与资料若有不实之处，本人承担一切相关责任。

本人签名：\_\_\_\_\_ 日期：\_\_\_\_\_

## 关于论文使用授权的说明

本人完全了解并同意北京邮电大学有关保留、使用学位论文的规定，即：北京邮电大学拥有以下关于学位论文的无偿使用权，具体包括：学校有权保留并向国家有关部门或机构送交学位论文，有权允许学位论文被查阅和借阅；学校可以公布学位论文的全部或部分内容，有权允许采用影印、缩印或其它复制手段保存。汇编学位论文，将学位论文的全部或部分内容编入有关数据库进行检索。（保密的学位论文在解密后遵守此规定）

本人签名：\_\_\_\_\_ 日期：\_\_\_\_\_

导师签名：\_\_\_\_\_ 日期：\_\_\_\_\_

# 基于 LLM 的交互式多模态图像编辑系统的设计 与搭建

## 摘要

这是中文摘要的部分。

它可以拥有多段。这是中文摘要的部分。

它可以拥有多段。

如果你写的太长，甚至可以到第二页。

**关键词** 北京邮电大学 本科生 毕业设计 模板 示例

# **Design and Construction of Interactive Multimodal Image Editing System Based on LLM**

## **ABSTRACT**

This is ABSTRACT.

You can write more than one paragraph here.

If your abstract is too long, it will take up more pages.

**KEY WORDS** BUPT undergraduate thesis template example

# 目 录

<b>第一章 绪论</b>	1
1.1 项目背景	1
1.2 项目意义	2
1.3 项目内容	2
<b>第二章 总体方案设计</b>	4
2.1 GUI	4
2.2 middleware	4
2.3 Stable Diffusion	5
2.4 OpenAI	6
2.5 ChatGLM2-6B	6
<b>第三章 GUI 的构建</b>	7
3.1 图像自动遮罩与优化	7
3.1.1 图像自动遮罩	7
3.1.2 对自动生成的遮罩进行优化	9
3.2 多模态	10
3.2.1 JSON 指令生成	10
3.2.2 JSON 指令校验	11
3.2.3 多指令处理	12
3.2.4 图像模型请求参数生成	12
3.3 图像修改建议	12
<b>第四章 middleware 的构建</b>	14
4.1 对多个平台的 API 进行配置和整合	14
4.2 使用 beego 框架提供 API 服务	14
<b>第五章 LLM 的微调</b>	15
5.1 LLM 微调数据集生成与性能评估方法	15
5.1.1 微调数据集生成	15
5.1.2 LLM 指令生成任性能评估方法	16
5.2 ChatGLM2-6B 针对指令生成任务的微调	16
5.3 各个 LLM 在本任务下的性能评估	16
<b>第六章 Stable Diffusion 及扩展的使用</b>	17
6.1 Stable Diffusion API 的使用	17
6.2 ControlNet 的使用及效果	17

---

6.3 roop 的使用及效果 .....	17
<b>第七章 系统实现效果与使用 .....</b>	<b>18</b>
7.1 系统实现效果 .....	18
7.2 系统使用方法 .....	18
<b>第八章 项目管理与维护 .....</b>	<b>19</b>
8.1 代码管理 .....	19
8.2 自动化测试 .....	19
8.3 持续集成与持续部署 .....	19
<b>第九章 基础模块示例 .....</b>	<b>20</b>
9.1 特殊文本类型 .....	20
9.1.1 脚注 .....	20
9.1.2 定义、定理与引理等 .....	20
9.1.3 中英文文献、学位论文引用 .....	20
9.2 图表及其引用 .....	21
9.3 公式与算法表示 .....	22
9.3.1 例子：基于主成分分析 .....	22
9.3.1.1 主成分分析算法 .....	22
9.3.1.2 主成分分析可信度评估方法 .....	25
9.4 代码表示 .....	25
9.4.1 直接书写代码在.tex 中 .....	25
9.4.2 引用代码文件 .....	25
9.5 列表样式 .....	26
9.5.1 使用圆点作为项目符号 .....	26
9.5.2 使用数字作为项目符号 .....	26
9.5.3 句中数字编号列表样式 .....	26
<b>第十章 为了目录撑到第二页 .....</b>	<b>27</b>
10.1 我不得不再添加一点内容 .....	27
10.2 尽管这些章节一点正文都没有 .....	27
10.3 是的 .....	27
10.4 真的没有 .....	27
10.5 我已经不知道说什么了 .....	27
10.6 如果有，我们就祝愿一下学校教务处什么时候转变一下思维 .....	27
10.7 把控制格式这种事情往前做 .....	27
10.8 不要总是觉得折磨学生是合理的 .....	27
10.9 你拿着教学管理岗位的工资 .....	27

10.10	你需要折磨一下你自己才对	27
10.11	不要觉得我对别人要求太高，对自己太低	27
10.12	我对自己要求低的话也不至于想要修订这份模板	27

参考文献

致 谢

附 录

外 文 资 料

外 文 译 文

开 题 报 告

中 期 检 查 表

教师指导毕业设计(论文)记录表

# 第一章 绪论

## 1.1 项目背景

随着技术的迅速发展，图像生成技术在多个行业中发挥着越来越重要的作用，尤其是在媒体娱乐、数字营销和智能医疗等领域。然而，尽管其应用广泛，传统的图像编辑技术仍面临许多挑战，尤其是在交互性和生成图像质量上的限制。

传统图像编辑工具往往依赖于专业的技术知识和复杂的操作界面，这对于普通用户来说是一个较大的门槛。用户需要花费大量时间学习如何使用这些工具，这限制了工具的普及性和易用性。这些工具的交互性通常较差，不能很好地根据用户的具体需求进行灵活调整和响应。

为了克服这些挑战，人们正在探索利用深度学习技术来改善图像编辑系统。深度学习，特别是扩散模型，已经在图像生成领域显示出了巨大的潜力。这些技术能够学习大量图像数据，自动提取复杂的特征，并生成高质量的图像。此外，结合最新的 GPT4 等性能优异的大语言模型，可以进一步提升系统的交互性，实现更加自然和直观的用户界面。

图像生成模型是深度学习领域中一个活跃的研究方向，主要致力于通过机器学习技术生成高质量的图像。这些模型在多种应用场景中均有广泛应用，包括艺术创作、游戏开发、电影制作等。目前，图像生成模型主要包括生成对抗网络 (GANs)、变分自编码器 (VAEs) 和扩散模型等。

扩散模型是近年来发展起来的一种新型生成模型，与传统的生成对抗网络 (GANs) 和变分自编码器 (VAEs) 相比，扩散模型在生成图像的质量和多样性方面展现出了卓越的性能。其基本原理是模拟从高质量数据分布到高熵噪声分布的逐步转变过程，然后再逆向这一过程以生成新的数据。扩散模型的工作流程可以分为两个阶段：前向扩散过程和反向生成过程。在前向扩散过程中，模型逐渐将数据中的信息转化为噪声，这一过程通常通过向数据中逐步加入高斯噪声来实现。在反向生成过程中，模型则需要学习如何从噪声状态恢复出原始数据的分布，这一过程通常通过训练一个参数化的神经网络来完成，网络的目标是最小化原始数据与生成数据之间的差异。扩散模型的关键优势在于其生成的图像具有较高的质量和自然性，这是因为模型在生成过程中能够更好地控制噪声的去除过程，并逐步精细化图像的细节。此外，扩散模型在训练过程中相对稳定，不易出现生成对抗网络中常见的模式崩溃问题。

大语言模型是人工智能领域中的一项核心技术，主要用于处理和理解自然语言。这些模型通过学习大量文本数据，能够生成文本、回答问题、翻译语言等。随着算力的提升和语料的增加，大语言模型已经取得了显著的进步，并在多个应用场景中展现出了巨大的潜力。近几年，随着深度学习技术的发展，尤其是 Transformers 架构的提出，大语言模型的性能得到了质的飞跃。Transformers 架构能够有效处理长距离依赖问题，并在许多自然语言处理任务中设定了新的性能标准，GPT (Generative Pre-trained Transformer)、

BERT (Bidirectional Encoder Representations from Transformers) 等主流模型，通过大规模的语料进行预训练，对语言的深层次结构和语义的理解能力有了显著的提升，在应用方面展现出广泛的适用性，包括但不限于文本生成、对话系统、语言翻译、内容摘要和情感分析等。

通过图像生成模型与大语言模型的结合，可以创建一个更加灵活且易用性强的图像编辑系统。这样的系统不仅能够提供更加直观的编辑界面，降低用户的操作难度，还能够根据用户的描述自动生成或修改图像，极大地提升生成图像的质量和编辑效率。例如，用户可以通过简单的语言指令，如“增加图片亮度”或“改变背景为海滩”，直接与编辑系统交互，系统能够理解这些指令并即时作出响应。

通过整合深度学习和语言模型技术，我们有望构建出一个全新的交互式图像编辑系统。这种系统不仅能够提供更高质量的图像编辑结果，更能够提供给用户更加自然的交互方式，为该领域的专业人士和普通用户都带来更多的可能性。

## 1.2 项目意义

通过融合先进的图像生成技术与大语言模型，本项目希望能搭建一个基于 LLM 的交互式图像编辑系统，提供更为易用、精准且高效的图像编辑工具，提升媒体娱乐、数字营销以及智能医疗等多个行业的图像处理能力。通过实现更加智能化和用户友好型的编辑系统，为广大用户带来前所未有的图像编辑体验，并为促进技术创新做出贡献。这样的研究与开发，为图像编辑技术的未来提供了一种可能性和一条新的探索和实践路径。

## 1.3 项目内容

该项目是一个多模态交互式图像编辑系统，它主要实现了 GUI、middleware、以及对 Stable Diffusion 和 ChatGLM2-6B 模型的修改与适配。在整体结构上，GUI、middleware、以及模型修改之间的相互关系和数据流向见图 1-1。

项目基于 Stable Diffusion 的开源项目 stable-diffusion-webui 进行了扩展，增强了其功能。系统可以调用不同的 Stable Diffusion 模型并结合多个扩展后的功能来对图像进行精细的修改和调整。这一功能的实现大大增强了系统对图像处理的灵活性和多样性。

项目还包括了对 ChatGLM2-6B 模型的微调。通过利用专门为本项目的需求生成的微调数据集进行微调，进一步提升了语言模型处理特定任务的能力和准确性，且能够利用 API 调用这些经过微调的 ChatGLM2-6B 模型。

通过调用 OpenAI 的 API，本项目实现了多项功能：利用 GPT4V 生成关于图像修改的建议，使用 GPT3.5Turbo 来辅助生成用于微调大语言模型的数据集以及图像修改指令，以及在 Stable Diffusion 模型不可用时，使用 DALL-E2 作为替代模型来进行图像修改。这些功能为基于 LLM 的交互式图像编辑系统提供了强大的工具。

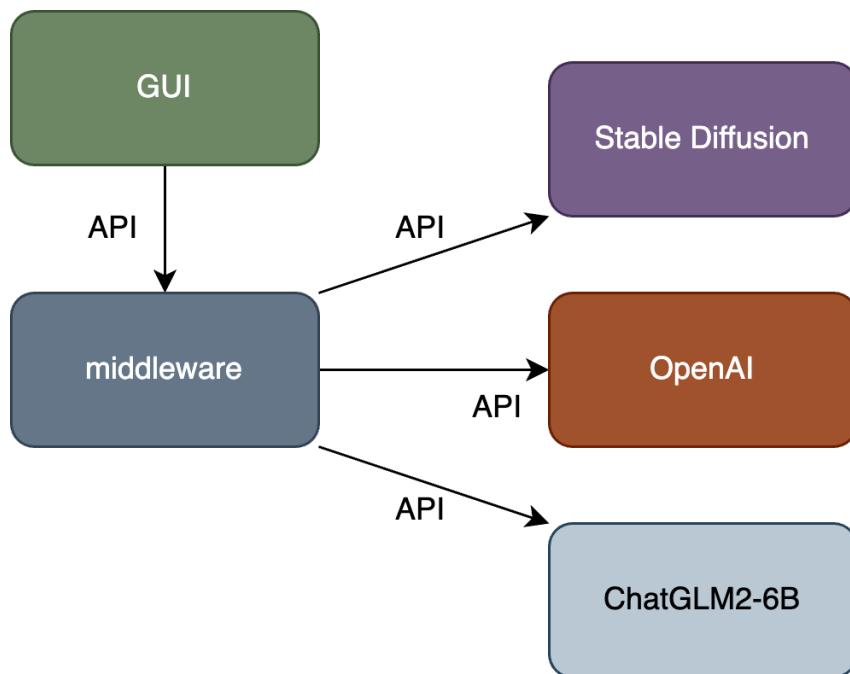


图 1-1

在 GUI 方面，本系统主要通过 Python 语言实现，构建了一个直观且用户友好的交互界面。该界面简洁易用，通过 API 调用向 middleware 发出请求，有效减轻了用户在执行计算密集型任务时对硬件资源的需求，从而降低了用户侧的使用门槛。

middleware 部分使用 Golang 语言构建了一个后端服务。这个服务不仅接入了 Stable Diffusion、ChatGLM2-6B、OpenAI 的 API，还将这些 API 进行了有效的整合，向 GUI 提供了一致风格的 API 接口。这种设计不仅提升了 GUI 调用多方 API 的便利性，也通过统一的配置和管理，极大地增强了系统的可维护性和稳定性。

## 第二章 总体方案设计

### 2.1 GUI

GUI 虽然承担计算任务较少，但却是承载本项目结构与逻辑的关键部分。通过使用符合规则的指令作为中枢，GUI 打通了大语言模型和图像生成模型之间的壁垒，使基于 LLM 的创新交互式图像编辑系统成为可能。GUI 的模块构成如表 2-1：

表 2-1 GUI 模块

模块	描述
BaseImage	接受上传的原始图片并预览
EditedImage	预览修改后的图片
Operation Board	执行指令
Settings	对系统进行设置
Chat	与大语言模型交互的聊天界面
Edit Image	对图像进行自定义遮罩和换脸等操作
Auto	执行自动化操作
Manual	系统使用说明

用户首先上传需要修改的图片，然后可在 Chat 模块中选择不同的大语言模型进行交互并得到相应的指令，最后在 Operation Board 模块中选择指令执行或一键全部执行。如果对自动生成的遮罩不满意，可在 Edit Image 中对遮罩进行修改。

在 Auto 模块中，用户可通过选择多张图片批量生成满足微调大语言模型微调所需的数据。其会循环地从给定的图片集中随机选择图片继续分割，将分割后的结果和特定的 prompt 通过 GPT3.5Turbo 生成对应的修改建议，再将分割的结果、生成的建议通过 GPT3.5Turbo 生成指令。

### 2.2 middleware

middleware 是项目的核心组件，通过整合多个平台的 API，为 GUI 提供统一的、简单易接入的 API 服务。其主要特点包括但不限于：1. API 整合：middleware 整合了多个平台的 API，包括图像生成模型、语言模型等，使得 GUI 可以通过统一的接口调用不同功能模块；2. 统一风格：middleware 设计了统一的 API 风格和路由规范，使得 GUI 可以轻松使用 API 服务，提高开发效率；3. 简单易接入：middleware 提供了简单易用的 API 服务，GUI 无需关注具体实现细节，只需按照简单的请求规范调用 API 即可；4. 稳定性和可靠性：middleware 基于 Golang 语言实现，具有高效的并发处理能力和稳定的运行性能，保证了 API 服务的稳定性和可靠性；5. 易于维护：middleware 采用了 Beego

框架，具有清晰的代码结构和模块化设计，易于维护和扩展，保证了项目的长期可持续发展。其向 GUI 提供的主要 API 如表 2-2 所示：

表 2-2 middleware 主要 API

API	路由	描述
PostSDTxt2Img	/v1/pics/txt2img	通过 Stable Diffusion 模型生成图片
PostSDImg2Img	/v1/pics/img2img	通过 Stable Diffusion 模型修改图片
PostDALLE2Edit	/v1/pics/openai/img2img	通过 DALL-E2 模型修改图片
GetLoras	/v1/pics/loras	获取可用的 LoRa 模型列表
PostHuggingFaceImgSegment	/v1/pics/huggingface/segment	获取图像分割结果
PostGPT3Dot5Turbo	/v1/chat/gpt3dot5turbo	调用 GPT3.5Turbo
PostGPT4	/v1/chat/gpt4	调用 GPT4
PostChatGLM2_6B	/v1/chat/glm2_6b	调用 ChatGLM2-6B
PostGPT4V	/v1/chat/gpt4v	调用 GPT4V

## 2.3 Stable Diffusion

Stable Diffusion 是一种基于扩散模型的深度学习图像生成模型，它能够根据文本描述生成高质量的图像。这个模型采用了条件生成技术，允许用户通过文本指令来引导图像的生成，使其在艺术创作、媒体娱乐、广告和数字营销等多个领域具有广泛的应用。其核心在于从噪声逆向还原生成高质量的图片。

sd-webui-controlnet<sup>1</sup>使用了 ControlNet<sup>[1]</sup> 的原理，旨在增强原有 Stable Diffusion 模型的图像生成控制能力。通过集成一个额外的控制网络（ControlNet），允许用户精确指导图像的具体内容，显著提升了生成图像的细节质量和一致性。

sd-webui-roop<sup>2</sup>基于 DeepFake<sup>[2]</sup>，允许用户在图片中进行面部替换，简化了面部交换的过程，无需训练特定的模型，大大降低了使用复杂度。

由于本项目对于图像生成模型的要求较高且需求复杂，为了便于结合 Stable Diffusion 模型和其他前沿研究成果及开源社区项目，本项目在构建 Stable Diffusion 模块时以开源项目 stable-diffusion-webui<sup>3</sup>为基础，结合 sd-webui-controlnet 和 sd-webui-roop 等扩展，通过 API 为 middleware 提供服务。

<sup>1</sup><https://github.com/Mikubill/sd-webui-controlnet>

<sup>2</sup><https://github.com/s0md3v/sd-webui-roop>

<sup>3</sup><https://github.com/AUTOMATIC1111/stable-diffusion-webui>

## 2.4 OpenAI

本项目使用了 OpenAI<sup>1</sup>的 GPT3.5Turo、GPT4、GPT4V、DALL-E2 等模型，通过 API 调用 OpenAI 的模型。

GPT-3.5 Turbo 是 OpenAI 开发的一款先进的自然语言处理模型，属于 GPT-3 系列的增强版本。这个模型在处理大量文本和生成文本方面表现出色，适用于聊天机器人、内容创作、文本摘要等应用。GPT-3.5 Turbo 优化了处理速度和响应时间，提高了交互效率。

GPT-4 是 GPT-3 的后续版本，代表了最新一代的语言预测和生成技术。它在模型结构和训练数据量上进行了大幅扩展，使其能够更准确地理解和生成复杂的文本。GPT-4 在理解上下文、维持一致性以及生成更自然的语言方面具有显著优势。

GPT-4V 是 GPT-4 的一个特殊版本，专门优化用于视觉任务，比如图像标注、视觉问答等。这个版本结合了文本和视觉处理能力，能够更好地理解和生成与图像相关的内容。

DALL-E 2 是一个先进的图像生成模型，专门设计用来创建新颖的图像和艺术作品。它可以根据用户提供的文本描述生成详细、高质量的图像。DALL-E 2 的核心优势在于其创造力和多样性，能够在遵循描述的同时，创造出独特和富有创意的视觉内容。

## 2.5 ChatGLM2-6B

ChatGLM2-6B 是由清华大学开发的第二代开源双语(中英)对话模型，基于 ChatGLM-6B 的成功经验并引入了多项新特性和性能提升。这款模型经过大规模预训练，实现了显著的性能提升，并在多个数据集上表现出色。ChatGLM2-6B 支持更长的对话上下文，并提高了推理速度和降低了显存占用，使得即使在资源有限的环境下也能有效运行。

本项目使用 ChatGLM2-6B 模型，使用开源项目 LLaMA-Factory<sup>2</sup>，利用本项目提供的数据自动生成功能所生成的数据集，使用 LoRa<sup>[3]</sup> 方法对模型进行微调以在本项目所需的任务中获得更佳的表现。微调后的模型通过 fastapi 提供 API 服务。

---

<sup>1</sup><https://openai.com>

<sup>2</sup><https://github.com/hiyouga/LLaMA-Factory>

## 第三章 GUI 的构建

在本项目中，GUI 作为关键组成部分，虽仅承担极少的计算任务，却在结构与逻辑上起着至关重要的作用。它通过使用标准化的指令连接大语言模型和图像生成模型，实现了基于 LLM 的创新交互式图像编辑系统。用户首先上传原始图片至 BaseImage 模块进行预览，之后可通过 Chat 模块与大语言模型进行交互，获取编辑指令。用户可以在 Operation Board 模块中选择单独或批量执行这些指令。若对自动生成的遮罩不满意，可在 Edit Image 模块中手动调整。此外，Auto 模块允许用户批量处理多张图片，自动生成数据以微调大语言模型。该过程包括图片的自动分割、利用 GPT3.5Turbo 生成编辑建议及相应的执行指令。这样的设计不仅提升了系统的效率，也优化了用户的交互体验。

### 3.1 图像自动遮罩与优化

由于本项目需要提供对图像进行部分修改的功能，所以需要在使用图像生成模型进行图像编辑时需要提供一个遮罩以明确需要修改的部分和不需要修改的部分。为了自动生成符合要求的遮罩，本项目借助图像分割和大语言模型的辅助，可通过两种方式生成自动遮罩：基于关键词对自动生成遮罩和基于已给出的点自动填充生成遮罩。两种方法都会首先使用图像分割模型对图像进行分割（如图 3-1），然后根据给出的要求对相应的部分进行遮罩生成原始的遮罩。受制于图像分割模型在边缘上的表现并不理想，需要对特定的分割部分进行处理以提高遮罩的质量，因此最后会通过本项目设计的优化算法生成最终的遮罩。



图 3-1 图像分割结果：(a)原始图像，(b)分割结果

#### 3.1.1 图像自动遮罩

基于关键词自动生成遮罩的方法会根据关键词和图像分割结果生成自动原始的遮罩，该功能会遍历每个给出的关键词，若关键词与分割结果之一吻合，则会对相应的分割区域进行遮罩，生成原始的遮罩（如图 3-2）。

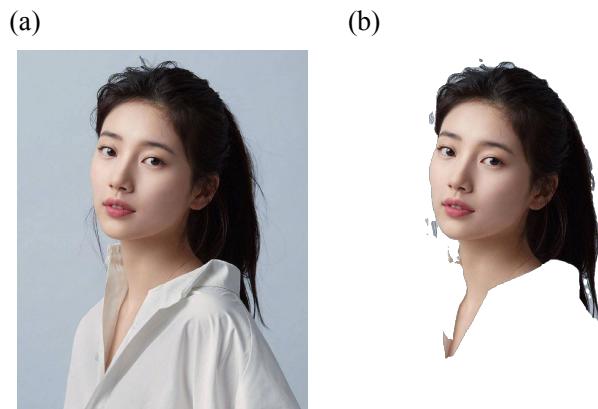


图 3-2 关键词自动生成遮罩结果: (a)原始图像, (b)keywords=[’Background’, ’Upper-clothes’, ’Dress’, ’Right-arm’] 得到的遮罩

基于已给出的点自动填充生成遮罩的方法会根据在图片中标记的点和图像分割结果生成自动原始的遮罩, 该功能会遍历每个给出的点, 将该点所在的部分全部进行遮罩, 最后生成原始的遮罩 (如图 3-3)。



图 3-3 基于已给出的点自动填充生成遮罩: (a)标记后的图像, (b)生成的遮罩

### 3.1.2 对自动生成的遮罩进行优化

由于分割模型性能的限制，生成的原始遮罩可能在某些细节上表现不佳而影响图像编辑模型的结果，因此设计了一个算法对自动生成的遮罩进行优化。该算法可以根据配置文件的设置，对特定的未被遮罩的部分在遮罩的边缘进行收缩。

---

#### 算法 1 遮罩优化算法

---

**输入：** 原始遮罩  $OriginMask$ ，图像分割结果  $SegmentResult$ ，配置文件  $Config$

**输出：** 优化后的遮罩  $OptimizedMask$

- 1: 获取遮罩与非遮罩的描边得到像素  $EdgePixels$
  - 2: 从配置文件和图像分割结果获取  $ConfigPixels$
  - 3: 仅保留出现在  $EdgePixels$  中的  $ConfigPixels$
  - 4: **for**  $pixcel$  in  $ConfigPixels$  **do**
  - 5:     Apply MinFilter Kernel(in  $Config$ ) in  $OriginMask[pixcel]$
  - 6: **end for**
  - 7: 得到优化后的遮罩  $OptimizedMask$
- 

由于该算法仅会对遮罩边缘上的像素进行卷积且在设计时充分考虑到了内存中像素的存储顺序的原因，虽然需要复杂的处理过程，但经过多次的迭代后算法的时间复杂度降低至  $O(mnr)$  ( $m, n$  表示图片的长宽， $r$  表示设置的优化强度)。算法实现的效果如图 3-4 所示，可见在发丝附近遮罩的质量得到了明显的改善。

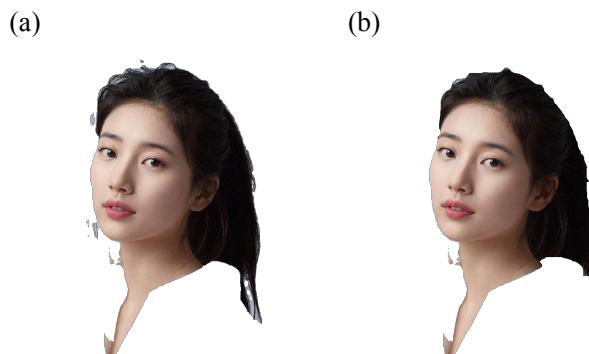


图 3-4 遮罩优化结果：(a)原始遮罩，(b)算法优化后的遮罩

## 3.2 多模态

如何打通大语言模型和图像生成模型是本项目的关键。本项目通过特定的 prompt 和图像分割结果，使用大语言模型生成 JSON 格式的指令并校验，并支持多轮对话。用户可有选择性地执行生成的指令或执行全部指令。系统首先会按照给定的规则对指令进行预处理和排序，然后通过指令生成请求参数来调用图像生成模型。多模态任务的实现方式如图 3-5。

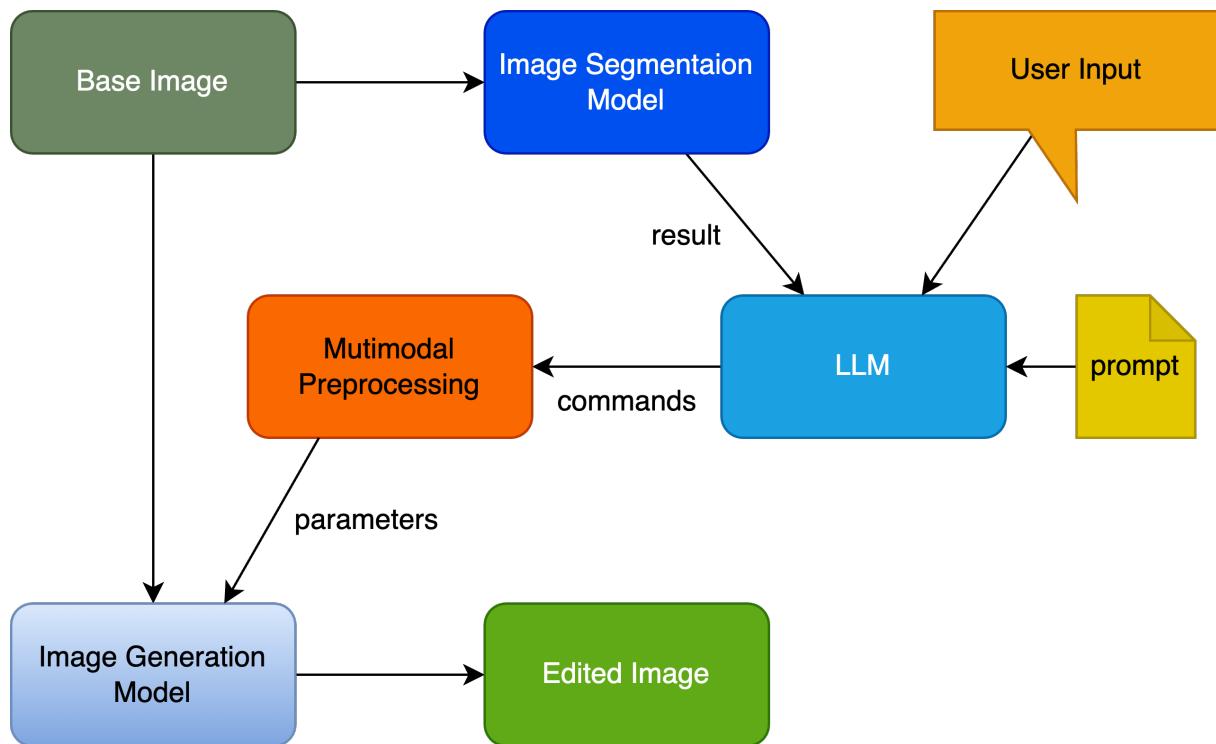


图 3-5

### 3.2.1 JSON 指令生成

JSON (JavaScript Object Notation) 是一种轻量级的数据交换格式，易于人阅读和编写，同时也易于机器解析和生成。它基于 JavaScript 的一个子集，但独立于语言，被广泛应用于许多编程语言中。JSON 主要用于网络应用间浏览器与服务器之间的数据传输。在 JSON 中，数据以键值对的形式存在，可以表示数组、布尔值、数字、对象或字符串。由于其简洁性和易于交互的特性，JSON 已成为 Web 应用中数据交换的主流技术。由于 JSON 应用范围广且大语言模型 JSON 处理能力较强，本项目采用此格式来承载大语言模型和图像生成模型的联系。

通过特定的 prompt 和图像分割结果以及用户输入的修改意图，本项目可使用 GPT3.5Turbo、GPT4、微调后的 ChatGLM2-6B 生成 JSON 指令。例：当图像分割结果为 *Background, Hair, Upper-clothes, Dress, Face, Right-arm*，用户输入为“将背景更换为蓝天白云，将衣服更改为白色的 T-shirt”时，生成的 JSON 指令如代码 3-1 所示。

代码 3-1 生成的指令

```

1  [
2      {
3          "command" : "change",
4          "paras" : [ ["Background", "Upper-clothes"] , "blue\u2022sky,\u2022white\u2022T-shirt"]
5      }
6  ]

```

### 3.2.2 JSON 指令校验

由于大语言模型生成指令不稳定，需要对生成指令的合规性进行校验。校验规则存储为一个 JSON 文件，以修改非面部和面部的指令为例，校验规则如代码 3-2 所示：

代码 3-2 指令校验规则

```

1  {
2      "face": {
3          "paras_type": [
4              "<class\u2022'str'\u2022"
5          ],
6          "paras_enum": null,
7          "paras_min": null,
8          "paras_max": null,
9          "combine": true,
10         "priority": 1
11     },
12     "change": {
13         "paras_type": [
14             "<class\u2022'list'\u2022",
15             "<class\u2022'str'\u2022"
16         ],
17         "paras_enum": null,
18         "paras_min": null,
19         "paras_max": null,
20         "combine": true,
21         "priority": 1
22     }
23 }

```

指令校验的算法如算法-2 所示。

---

### 算法 2 JSON 指令校验算法

---

输入：待校验的指令 *JsonCommands*、校验规则 *Rules*

输出：校验后的指令 *ValidJsonCommands*

```

1: for Command in JsonCommands do
2:   if Command satisfy Rules then
3:     Add Command to ValidJsonCommands
4:   end if
5: end for

```

---

### 3.2.3 多指令处理

由于一次执行可能会涉及到多个指令，会遇到指令重复、指令优先性等问题，所以会对需要执行的指令进行合并与排序。指令合并的算法如下：

---

### 算法 3 多指令处理算法

---

输入：待处理的指令 *OriginCommands*、指令合并规则 *Rules*、指令优先性 *Priority*

输出：处理后的指令 *ProcessedCommands*

```

1: for Command in OriginCommands do
2:   if Same Command type already in ProcessedCommands then
3:     Combine Command to the same one in ValidJsonCommands use Rules
4:   end if
5: end for
6: 根据 Priority 对 ValidJsonCommands 进行排序

```

---

### 3.2.4 图像模型请求参数生成

图像模型请求参数生成较为复杂，对于某个参数，其可能来源于 GUI 中可修改的设置，可能来源于指令，可能来源于模版，否则设置为默认参数。由于每个参数来说，其来源的优先性可能不一致，因此设计了算法-4 来生成图像模型请求参数。

## 3.3 图像修改建议

本项目提供了根据图像自动生成图像建议的功能。由于传统的大语言模型只能接受文本输入，因此本项目采用了 GPT4V 来自动生成图像修改建议。

GPT-4V 是由 OpenAI 开发的多模态大型语言模型，是 GPT 系列基础模型的第四代。该模型具有视觉能力，可以将图片作为输入，进行各种任务，例如描述图片中的幽默、总结截图文本、回答包含图表的考试题目等。

用户通过 GUI 界面的 Advise 按键，可以生成建议并将其转换为指令。

---

#### 算法 4 图像模型请求参数生成算法

---

输入： 指令 *Command*、设置 *Settings*、默认参数 *Default*、参数来源优先性 *PriorityRules*

输出： 图像模型请求参数 *Parameters*

```
1: for ParaKey in Parameters do
2:   Get Template from Command
3:   if ParaKey found in Command or Settings or Template then
4:     Choose the highest priority source use PriorityRules[ParaKey]
5:   else
6:     Set this parameter to Default[ParaKey]
7:   end if
8: end for
```

---

## 第四章 middleware 的构建

在本项目中, middleware 作为核心组件, 通过整合来自不同平台的 API, 为 GUI 提供了统一且易于接入的 API 服务。其整合了多个图像生成和大语言语言模型, 通过统一的接口设计, 使得 GUI 能够方便地调用所需的功能。此外, middleware 采用了 Golang 语言和 Beego 框架, 不仅保证了 API 服务的高并发处理能力和稳定性, 还通过模块化的设计提高了系统的可维护性和可扩展性。主要 API 服务包括 Stable Diffusion 和 DALL-E2 模型、图像分割模型, 以及多种大语言模型。这样的架构设计不仅优化了开发效率, 也确保了系统的稳定运行和长期发展。

### 4.1 对多个平台的 API 进行配置和整合

middleware 通过整合不同平台如图像生成模型、语言模型等的 API, 使得 GUI 开发者可以通过一个统一的接口调用多种功能。这种整合不仅包括 API 的聚合, 还涉及统一 API 调用的风格和路由规范, 不仅保证了 API 服务的高稳定性和可靠性, 还便于日后的维护和扩展。例如, 无论是调用 ChatGLM2-6B 模型还是多种 GPT 模型, GUI 都能通过相同的结构化请求方式访问不同的服务。

### 4.2 使用 beego 框架提供 API 服务

middleware 组件提供了多个 API 服务, 以便 GUI 可以高效地使用各种模型和工具。*PostSDTxt2Img* 和 *PostSDImg2Img* 是通过 Stable Diffusion 模型来生成或修改图像的 API, 这使得用户能够通过简单的 API 调用, 进行复杂的图像生成和编辑操作。*GetLoras* 这个 API 用于获取 Stable Diffusion 模型可用的 LoRa 模型列表。*PostDALLE2Edit* 提供了利用 DALL-E2 模型修改图片的功能, 这进一步扩展了图像处理的能力, 在 Stable Diffusion 模型不可用时作为替代。*PostHuggingFaceImgSegment* API 通过可在部署在本地的分割模型不可用时通过调用 Hugging Face 上的图像分割模型 API 来实现图像分割。在文本处理方面, *PostGPT3Dot5Turbo*、*PostGPT4* 和 *PostGPT4V* 等 API 利用 OpenAI 的不同版本 GPT 模型来处理指令理解和生成任务, *PostChatGLM2\_6B* 可调用微调后的 ChatGLM2-6B 模型来进行指令生成。

## 第五章 LLM 的微调

大语言模型（LLM）的微调是一种调整预训练模型以更好地适应特定任务或应用场景的过程。原始的 ChatGLM2-6B 模型已经通过大规模数据预训练，具备了强大的语言理解和生成能力，通过针对特定任务的微调，可以进一步提升模型在指令生成任务的表现。本项目利用开源项目 LLaMA-Factory<sup>1</sup>和本项目中自动生成的指令生成任务数据集对 ChatGLM2-6B 模型进行微调。

### 5.1 LLM 微调数据集生成与性能评估方法

#### 5.1.1 微调数据集生成

由于没有适用于本任务的开源数据集，本项目尝试建立一个自动化工作流程，通过利用 ATR Dataset<sup>2</sup>调用多个模型和一定的校验规则来生成所需的数据并整合为一个数据集。数据集生成的流程如图 5-1。同时，在系统使用过程中产生的数据可设置是否保存，这些数据也会在运行数据集生成脚本添加到数据集中。

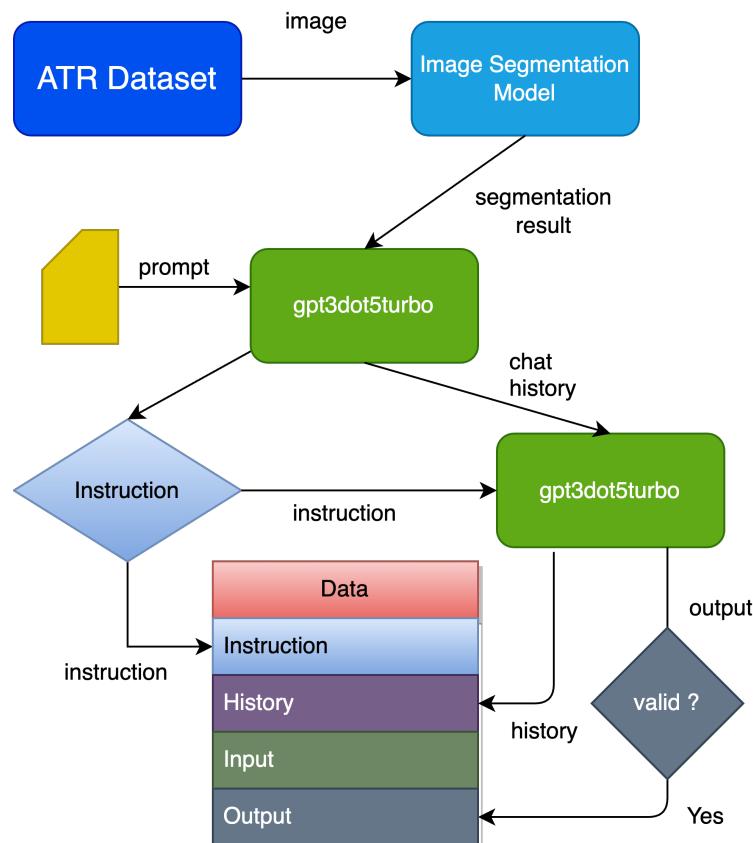


图 5-1

<sup>1</sup><https://github.com/hiyouga/LLaMA-Factory>

<sup>2</sup><https://github.com/lemondan/HumanParsing-Dataset>

### 5.1.2 LLM 指令生成任性能评估方法

由于需要一种直观的性能评估方法来对不同的大语言模型在指令生成任务上的性能进行评估，本项目设计了算法对模型性能进行量化。

## 5.2 ChatGLM2-6B 针对指令生成任务的微调

## 5.3 各个 LLM 在本任务下的性能评估

## 第六章 Stable Diffusion 及扩展的使用

6.1 Stable Diffusion API 的使用

6.2 ControlNet 的使用及效果

6.3 roop 的使用及效果

## 第七章 系统实现效果与使用

### 7.1 系统实现效果

### 7.2 系统使用方法

## 第八章 项目管理与维护

8.1 代码管理

8.2 自动化测试

8.3 持续集成与持续部署

## 第九章 基础模块示例

### 9.1 特殊文本类型

#### 9.1.1 脚注

社交媒体是一种供用户创建在线社群来分享信息、观点、个人信息和其它内容（如视频）的电子化交流平台，社交网络服务（social network service, SNS）和微博客（microblogging）都属于社交媒体的范畴<sup>[4]</sup>，国外较为知名的有 Facebook<sup>1</sup>、Instagram<sup>2</sup>、Twitter<sup>3</sup>、LinkedIn<sup>4</sup>等，国内较为知名的有新浪微博<sup>5</sup>。

在社交媒体的强覆盖下，新闻信息的传播渠道也悄然发生了变化。<sup>[5]</sup>

#### 9.1.2 定义、定理与引理等

**定义 9.1** 这是一条我也不知道在说什么的定义，反正我就是写在这里做个样子罢了，也没人会仔细读。<sup>[6]</sup>

**定理 9.1** 这是一条我也不知道在说什么的定理，反正我就是写在这里做个样子罢了，也没人会仔细读。

**公理 9.1** 这是一条我也不知道在说什么的公理，反正我就是写在这里做个样子罢了，也没人会仔细读。

**引理 9.1** 这是一条我也不知道在说什么的引理，反正我就是写在这里做个样子罢了，也没人会仔细读。

**命题 9.1** 这是一条我也不知道在说什么的命题，反正我就是写在这里做个样子罢了，也没人会仔细读。

**推论 9.1** 这是一条我也不知道在说什么的推论，反正我就是写在这里做个样子罢了，也没人会仔细读。

#### 9.1.3 中英文文献、学位论文引用

根据美国皮尤研究中心的 2017 年 9 月发布的调查结果<sup>[7]</sup>，67% 的美国民众会从社交媒体上获取新闻信息，其中高使用频率用户占 20%。在国内，中国互联网信息中心

本项目来源于科研项目“基于 LATEX 的本科毕业设计”，项目编号 1124

<sup>1</sup><http://www.facebook.com/>

<sup>2</sup><https://www.instagram.com/>

<sup>3</sup><http://www.twitter.com/>

<sup>4</sup><http://www.linkedin.com/>

<sup>5</sup><http://www.weibo.com/>

《2016 年中国互联网新闻市场研究报告》<sup>[8]</sup>也显示，社交媒体已逐渐成为新闻获取、评论、转发、跳转的重要渠道，在 2016 年下半年，曾经通过社交媒体获取过新闻资讯的用户比例高达 90.7%，在微信、微博等社交媒体参与新闻评论的比例分别为 62.8% 和 50.2%。社交媒体正在成为网络上热门事件生成并发酵的源头，在形成传播影响力后带动传统媒体跟进报道，最终形成更大规模的舆论浪潮。<sup>[9]</sup>

在国内，新浪微博由于其发布方便、传播迅速、受众广泛且总量大的特点，成为了虚假信息传播的重灾区：《中国新媒体发展报告（2013）》<sup>[?]</sup>显示，2012 年的 100 件微博热点舆情案例中，有超过 1/3 出现谣言；《中国新媒体发展报告（2015）》<sup>[?]</sup>对 2014 年传播较广、比较典型的 92 条假新闻进行了多维度分析，发现有 59% 的虚假新闻首发于新浪微博。

此等信息的传播严重损害了有关公众人物的名誉权，降低了社交媒体服务商的商业美誉度，扰乱了网络空间秩序，冲击着网民的认知，极易对民众造成误导，带来诸多麻烦和经济损失，甚至会导致社会秩序的混乱。针对社交媒体谣言采取行动成为了有关部门、服务提供商和广大民众的共同选择。<sup>[6]</sup>

## 9.2 图表及其引用

此处引用了简单的表 9-1。

请注意，**LATEX** 的图表排版规则决定了图表不一定会乖乖呆在你插入的地方，这是为了避免 Word 中由于图片尺寸不匹配在页面下部出现的空白，所以请不要使用“下图”“下表”作为指向文字，应使用“图 1-1 所示”这样的表述。

表 9-1 基于浏览者行为的特征

特征	描述	形式与理论范围
点赞量	微博的点赞数量	数值，N
评论量	微博的评论数量	数值，N
转发量	微博的转发数量	数值，N

此处引用了复杂的表 9-2。

表 9-2 基于浏览者行为的复杂特征

类别	特征	不知道叫什么的表头	
		描述	形式与理论范围
正常互动	点赞量	微博的点赞数量	数值，N
	评论量	微博的评论数量	数值，N
	转发量	微博的转发数量	数值，N
非正常互动	羡慕量	微博的羡慕数量	数值，N

此处展示了更专业的表 9-3，一个好的表格没有竖线。

表 9-3 红警 2 名词解释

术语类别	缩略语	解释
	兵营	兵营 (Barracks)，《命令与征服 红色警戒 2：尤里的复仇》游戏中的一种生产建筑，用以生产步兵单位
	建造场	建造场 (Construction Yard)，《命令与征服 红色警戒 2：尤里的复仇》游戏中的一种基础建筑，用以支持其他建筑的建造
	矿厂	矿石精炼厂 (Ore Refinery)，《命令与征服 红色警戒 2：尤里的复仇》游戏中的一种资源建筑，用以将矿车采集的矿石转化为游戏资金
游戏	空指	空指部 (Airforce Command Headquarters)，《命令与征服 红色警戒 2：尤里的复仇》游戏中的一种资源建筑，用以提供雷达功能和 T2 科技及生产部分空军单位
	相机	游戏术语，特指游戏内的观察区域和视角
	重工	战车工厂 (War Factory)，《命令与征服 红色警戒 2：尤里的复仇》游戏中的一种生产建筑，用以生产载具单位
	战争迷雾	游戏术语，《命令与征服 红色警戒 2：尤里的复仇》中指黑色的未探索区域

此处引用了一张图。图 9-1 表示的是一个由含有 4 个神经元的输入层、含有 3 个神经元的隐藏层和含有 4 个神经元的输出层组成的自编码器，+1 代表偏置项。

## 9.3 公式与算法表示

### 9.3.1 例子：基于主成分分析

#### 9.3.1.1 主成分分析算法

下面对主成分分析进行介绍。

主成分分析是一种简单的机器学习算法，其功能可以从两方面解释：一方面可以认为它提供了一种压缩数据的方式，另一方面也可以认为它是一种学习数据表示的无监督学习算法。<sup>[?]</sup> 通过 PCA，我们可以得到一个恰当的超平面及一个投影矩阵，通过投影矩阵，样本点将被投影在这一超平面上，且满足最大可分性（投影后样本点的方差最大化），直观上讲，也就是能尽可能分开。

对中心化后的样本点集  $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_i, \dots, \mathbf{x}_m\}$  (有  $\sum_{i=1}^m \mathbf{x}_i = 0$ )，考虑将其最大可分地投影到新坐标系  $\mathbf{W} = \{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_i, \dots, \mathbf{w}_d\}$ ，其中  $\mathbf{w}_i$  是标准正交基向量，满足  $\|\mathbf{w}_i\|_2 = 1$ ,  $\mathbf{w}_i^T \mathbf{w}_j = 0$  ( $i \neq j$ )。假设我们需要  $d'$  ( $d' < d$ ) 个主成分，那么样

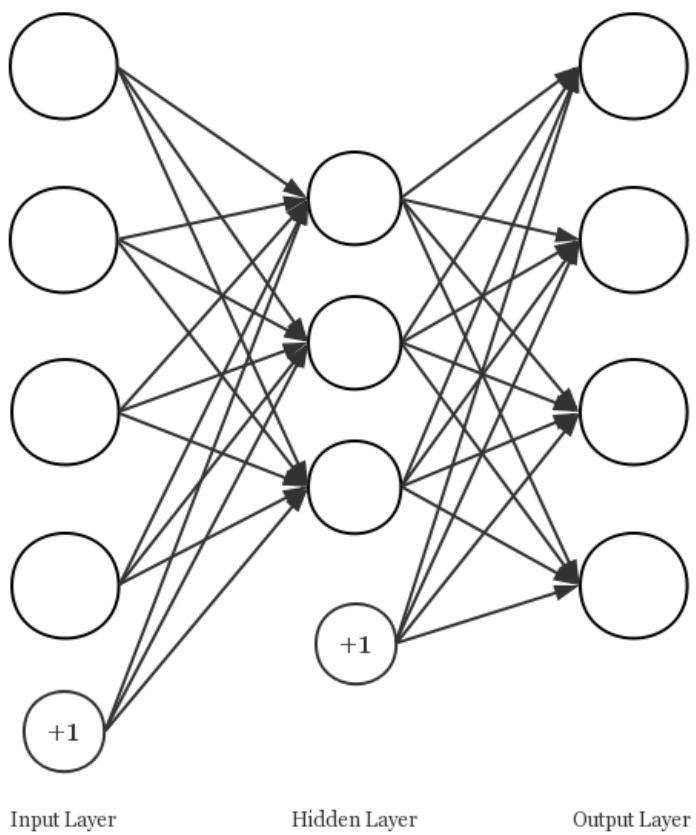


图 9-1 自编码器结构

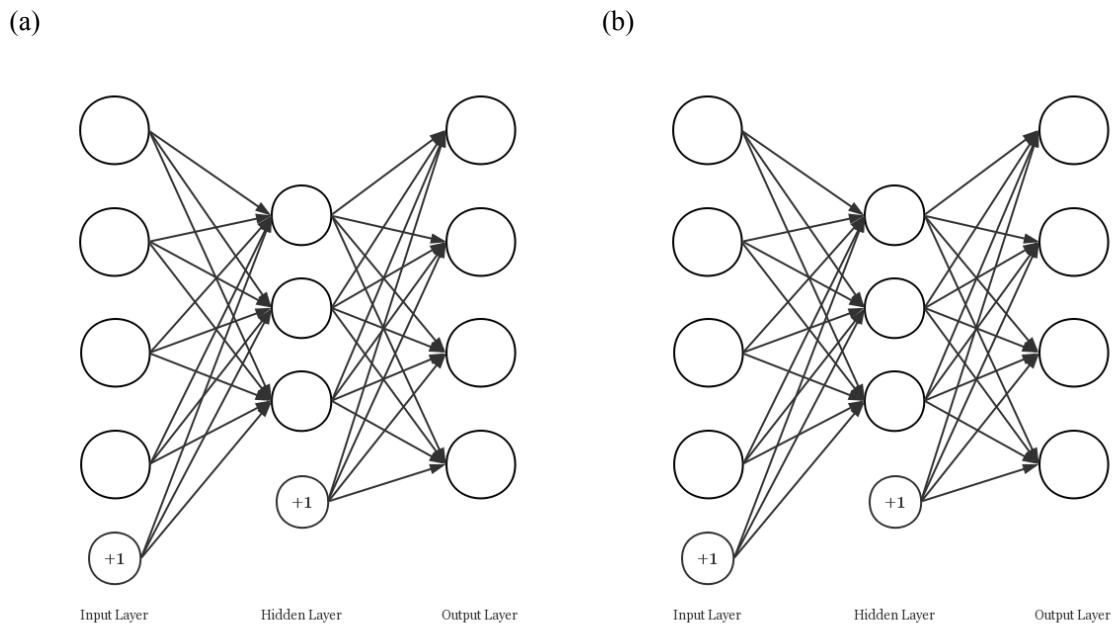


图 9-2 这是两个自编码器结构，我就是排一下子图的效果：(a)左边的自编码器，(b)右边的自编码器

本点  $\mathbf{x}_i$  在低维坐标系中的投影是  $\mathbf{z}_i = (z_{i1}; z_{i2}; \dots; z_{id'})$ , 其中  $z_{ij} = \mathbf{w}_j^\top \mathbf{x}_i$ , 是  $\mathbf{x}_i$  在低维坐标系下第  $j$  维的坐标。对整个样本集, 投影后样本点的方差是

$$\begin{aligned} & \frac{1}{m} \sum_{i=1}^m \mathbf{z}_i^\top \mathbf{z}_i \\ &= \frac{1}{m} \sum_{i=1}^m (\mathbf{x}_i^\top \mathbf{W})^\top (\mathbf{x}_i^\top \mathbf{W}) \\ &= \frac{1}{m} \sum_{i=1}^m \mathbf{W}^\top \mathbf{x}_i \mathbf{x}_i^\top \mathbf{W} \\ &= \frac{1}{m} \mathbf{W}^\top \mathbf{X} \mathbf{X}^\top \mathbf{W} \end{aligned} \quad \text{式 (9-1)}$$

由于我们知道新坐标系  $\mathbf{W}$  的列向量是标准正交基向量, 且样本点集  $\mathbf{X}$  已经过中心化, 则 PCA 的优化目标可以写为

$$\begin{aligned} \max_{\mathbf{W}} \quad & \text{tr}(\mathbf{W}^\top \mathbf{X} \mathbf{X}^\top \mathbf{W}) \\ \text{s.t.} \quad & \mathbf{W}^\top \mathbf{W} = \mathbf{I} \end{aligned} \quad \text{式 (9-2)}$$

由于  $\mathbf{X} \mathbf{X}^\top$  是协方差矩阵, 那么只需对它做特征值分解, 即

$$\mathbf{X}^\top \mathbf{X} = \mathbf{W} \Lambda \mathbf{W}^\top \quad \text{式 (9-3)}$$

其中  $\Lambda = \text{diag}(\boldsymbol{\lambda})$ ,  $\boldsymbol{\lambda} = \{\lambda_1, \lambda_2, \dots, \lambda_m\}$ 。

具体地, 考虑到它是半正定矩阵的二次型, 存在最大值, 可对式 (9-2) 使用拉格朗日乘数法

$$\mathbf{X} \mathbf{X}^\top \mathbf{w}_i = \lambda_i \mathbf{w}_i \quad \text{式 (9-4)}$$

之后将求得的特征值降序排列, 取前  $d'$  个特征值对应的特征向量组成所需的投影矩阵  $\mathbf{W}' = (\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_{d'})$ , 即可得到 PCA 的解。PCA 算法的描述如算法5所示。

---

#### 算法 5 主成分分析 (PCA)

---

输入: 样本集  $\mathbf{x} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_i, \dots, \mathbf{x}_m\}$ , 低维空间维数  $d'$

输出: 投影矩阵  $\mathbf{W}' = (\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_{d'})$

- 1: 对所有样本中心化  $\mathbf{x}_i \leftarrow \mathbf{x}_i - \frac{1}{m} \sum_{i=1}^m \mathbf{x}_i$
  - 2: 计算样本的协方差  $\mathbf{X} \mathbf{X}^\top$
  - 3: 对协方差矩阵  $\mathbf{X} \mathbf{X}^\top$  做特征值分解
  - 4: 取最大的  $d'$  个特征值所对应的特征向量  $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_{d'}$
-

### 9.3.1.2 主成分分析可信度评估方法

记待判定微博  $w_0$  的经典特征向量为  $f_0^c$ , 它的发布者在  $w_0$  前发布的  $k$  条微博为  $W = w_1, w_2, \dots, w_k$ , 这  $k$  条微博对应的经典特征向量集为  $F_W^c = \{f_1^c, f_2^c, \dots, f_k^c\}$ 。令  $label = 1$  代表谣言,  $label = 0$  代表非谣言。算法的具体流程如算法6所示。

---

#### 算法 6 基于 PCA 的信息可信度评估

---

输入:  $f_0^c, F_W^c$ , 保留主成分数  $n$

输出: 标签  $label \in \{0, 1\}$

- 1: 对所有特征向量应用 PCA, 保留前  $n$  个主成分  $\mathbf{o}_i^c \leftarrow PCA(f_i^c, n) \quad (i = 0, 1, \dots, k)$
  - 2: 计算  $F_W^c$  中各向量的平均距离  $\mu$  和标准差  $\sigma$
  - 3: 计算阈值  $thr = \mu/\sigma$
  - 4: **if**  $\min_{1 < j \leq k} \|\mathbf{o}_0^c - \mathbf{o}_j^c\|_2 > thr$  **then**
  - 5:      $label \leftarrow 1$
  - 6: **else**
  - 7:      $label \leftarrow 0$
  - 8: **end if**
- 

## 9.4 代码表示

### 9.4.1 直接书写代码在.tex 中

下面的代码9-1是用 Python 编写的加法函数。

代码 9-1 加法

```
1 def plusFunc(a, b):
2     return a + b
```

### 9.4.2 引用代码文件

下面的代码9-2是用 Python 文件中引入的倒序打印  $x$  到 1 的函数, 请查看 code 文件夹。

代码 9-2 倒序打印数字

```
1 def numbers(x):
2     if x > 0:
3         print(x)
4         numbers(x-1)
```

## 9.5 列表样式

### 9.5.1 使用圆点作为项目符号

- 第一章为基础模块示例，是的，本章的名字就是基础模块示例，正如你看到这个样子。
- 第二章为不存在，是的，其实它不存在。

### 9.5.2 使用数字作为项目符号

1. 第一章为基础模块示例，是的，本章的名字就是基础模块示例，正如你看到这个样子。
2. 第二章为不存在，是的，其实它不存在。

### 9.5.3 句中数字编号列表样式

1. 第一章为基础模块示例，是的，本章的名字就是基础模块示例，正如你看到这个样子；
2. 第二章为不存在，是的，其实它不存在。

## 第十章 为了目录撑到第二页

10.1 我不得不再添加一点内容

10.2 尽管这些章节一点正文都没有

10.3 是的

10.4 真的没有

10.5 我已经不知道说什么了

10.6 如果有，我们就祝愿一下学校教务处什么时候转变一下思维

10.7 把控制格式这种事情往前做

10.8 不要总是觉得折磨学生是合理的

10.9 你拿着教学管理岗位的工资

10.10 你需要折磨一下你自己才对

10.11 不要觉得我对别人要求太高，对自己太低

10.12 我对自己要求低的话也不至于想要修订这份模板

## 参考文献

- [1] Zhang Lvmin, Rao Anyi, Agrawala Maneesh. Adding conditional control to text-to-image diffusion models [C]. In Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023 : 3836–3847.
- [2] Van Huynh Nguyen, Hoang Dinh Thai, Nguyen Diep N et al. DeepFake: Deep dueling-based deception strategy to defeat reactive jammers [J]. IEEE Transactions on Wireless Communications. 20 (10). 2021: 6898–6914.
- [3] Hu Edward J, Shen Yelong, Wallis Phillip et al. Lora: Low-rank adaptation of large language models [J]. arXiv preprint arXiv:2106.09685. 2021.
- [4] Merriam-Webster. Social Media [EB/OL]. 2018 [2018-04-15]. <http://www.merriam-webster.com/dictionary/socialmedia>.
- [5] Vosoughi Soroush, Roy Deb, Aral Sinan. The spread of true and false news online [J]. Science. 359 (6380). 2018, 3: 1146–1151.
- [6] 周兴. 基于深度学习的谣言检测及模式挖掘 [学位论文]. 中国科学院大学, 2017.
- [7] Pew Research Center. News Use Across Social Media Platforms 2017 [EB/OL]. 2017 [2018-04-15]. <http://www.journalism.org/2017/09/07/news-use-across-social-media-platforms-2017>.
- [8] 中国互联网络信息中心. 2016 年中国互联网新闻市场研究报告 [EB/OL]. 2017 [2018-04-15]. <http://www.cnnic.cn/hlwfzyj/hlwzbg/mtbg/201701/P020170112309068736023.pdf>.
- [9] Yang Fan, Liu Yang, Yu Xiaohui et al. Automatic detection of rumor on Sina Weibo [C]. In Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics. New York, NY, USA. 2012 : 1–7.

## 致 谢

此处请写致谢的内容。

它可以有多段。

# 附录

## 附录 1 缩略语表

表 附-1 基于浏览者行为的特征

特征	描述	形式与理论范围
点赞量	微博的点赞数量	数值, $\mathbb{N}$
评论量	微博的评论数量	数值, $\mathbb{N}$
转发量	微博的转发数量	数值, $\mathbb{N}$

表 附-2 基于浏览者行为的复杂特征

类别	特征	不知道叫什么的表头	
		描述	形式与理论范围
正常互动	点赞量	微博的点赞数量	数值, $\mathbb{N}$
	评论量	微博的评论数量	数值, $\mathbb{N}$
	转发量	微博的转发数量	数值, $\mathbb{N}$
非正常互动	羡慕量	微博的羡慕数量	数值, $\mathbb{N}$

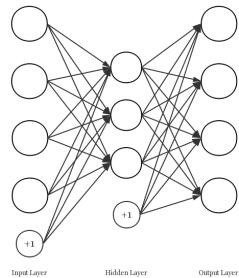


图 附-1 自编码器结构

代码 附-1 减法

```

1 def minusFunc(a, b):
2     return a - b

```

$$\max_{\mathbf{W}} \operatorname{tr}(\mathbf{W}^T \mathbf{X} \mathbf{X}^T \mathbf{W})$$

式 (附-1)

## 附录 2 数学符号

### 数和数组

$a$	标量 (整数或实数)
$\mathbf{a}$	向量
$dim()$	向量的维数
$A$	矩阵
$A^T$	矩阵 $A$ 的转置
$I$	单位矩阵 (维度依据上下文而定)
$diag(\mathbf{a})$	对角方阵, 其中对角元素由向量 $\mathbf{a}$ 确定

## SOCIAL SCIENCE

# The spread of true and false news online

Soroush Vosoughi,<sup>1</sup> Deb Roy,<sup>1</sup> Sinan Aral<sup>2\*</sup>

We investigated the differential diffusion of all of the verified true and false news stories distributed on Twitter from 2006 to 2017. The data comprise ~126,000 stories tweeted by ~3 million people more than 4.5 million times. We classified news as true or false using information from six independent fact-checking organizations that exhibited 95 to 98% agreement on the classifications. Falsehood diffused significantly farther, faster, deeper, and more broadly than the truth in all categories of information, and the effects were more pronounced for false political news than for false news about terrorism, natural disasters, science, urban legends, or financial information. We found that false news was more novel than true news, which suggests that people were more likely to share novel information. Whereas false stories inspired fear, disgust, and surprise in replies, true stories inspired anticipation, sadness, joy, and trust. Contrary to conventional wisdom, robots accelerated the spread of true and false news at the same rate, implying that false news spreads more than the truth because humans, not robots, are more likely to spread it.

**F**oundational theories of decision-making (1–3), cooperation (4), communication (5), and markets (6) all view some conceptualization of truth or accuracy as central to the functioning of nearly every human endeavor. Yet, both true and false information spreads rapidly through online media. Defining what is true and false has become a common political strategy, replacing debates based on a mutually agreed on set of facts. Our economies are not immune to the spread of falsity either. False rumors have affected stock prices and the motivation for large-scale investments, for example, wiping out \$130 billion in stock value after a false tweet claimed that Barack Obama was injured in an explosion (7). Indeed, our responses to everything from natural disasters (8, 9) to terrorist attacks (10) have been disrupted by the spread of false news online.

New social technologies, which facilitate rapid information sharing and large-scale information cascades, can enable the spread of misinformation (i.e., information that is inaccurate or misleading). But although more and more of our access to information and news is guided by these new technologies (11), we know little about their contribution to the spread of falsity online. Though considerable attention has been paid to anecdotal analyses of the spread of false news by the media (12), there are few large-scale empirical investigations of the diffusion of misinformation or its social origins. Studies of the spread of misinformation are currently limited to analyses of small, ad hoc samples that ignore two of the most important scientific questions: How do truth and falsity diffuse differently, and what factors of human judgment explain these differences?

Current work analyzes the spread of single rumors, like the discovery of the Higgs boson (13) or the Haitian earthquake of 2010 (14), and multiple rumors from a single disaster event, like the Boston Marathon bombing of 2013 (10), or it develops theoretical models of rumor diffusion (15), methods for rumor detection (16), credibility evaluation (17, 18), or interventions to curtail the spread of rumors (19). But almost no studies comprehensively evaluate differences in the spread of truth and falsity across topics or examine why false news may spread differently than the truth. For example, although Del Vicario *et al.* (20) and Bessi *et al.* (21) studied the spread of scientific and conspiracy-theory stories, they did not evaluate their veracity. Scientific and conspiracy-theory stories can both be either true or false, and they differ on stylistic dimensions that are important to their spread but orthogonal to their veracity. To understand the spread of false news, it is necessary to examine diffusion after differentiating true and false scientific stories and true and false conspiracy-theory stories and controlling for the topical and stylistic differences between the categories themselves. The only study to date that segments rumors by veracity is that of Friggeri *et al.* (19), who analyzed ~4000 rumors spreading on Facebook and focused more on how fact checking affects rumor propagation than on how falsity diffuses differently than the truth (22).

In our current political climate and in the academic literature, a fluid terminology has arisen around “fake news,” foreign interventions in U.S. politics through social media, and our understanding of what constitutes news, fake news, false news, rumors, rumor cascades, and other related terms. Although, at one time, it may have been appropriate to think of fake news as referring to the veracity of a news story, we now believe that this phrase has been irredeemably polarized in our current political and media climate. As politicians have implemented a political strategy of labeling news sources that do not

support their positions as unreliable or fake news, whereas sources that support their positions are labeled reliable or not fake, the term has lost all connection to the actual veracity of the information presented, rendering it meaningless for use in academic classification. We have therefore explicitly avoided the term fake news throughout this paper and instead use the more objectively verifiable terms “true” or “false” news. Although the terms fake news and misinformation also imply a willful distortion of the truth, we do not make any claims about the intent of the purveyors of the information in our analyses. We instead focus our attention on veracity and stories that have been verified as true or false.

We also purposefully adopt a broad definition of the term news. Rather than defining what constitutes news on the basis of the institutional source of the assertions in a story, we refer to any asserted claim made on Twitter as news (we defend this decision in the supplementary materials section on “reliable sources,” section S1.2). We define news as any story or claim with an assertion in it and a rumor as the social phenomena of a news story or claim spreading or diffusing through the Twitter network. That is, rumors are inherently social and involve the sharing of claims between people. News, on the other hand, is an assertion with claims, whether it is shared or not.

A rumor cascade begins on Twitter when a user makes an assertion about a topic in a tweet, which could include written text, photos, or links to articles online. Others then propagate the rumor by retweeting it. A rumor’s diffusion process can be characterized as having one or more cascades, which we define as instances of a rumor-spreading pattern that exhibit an unbroken retweet chain with a common, singular origin. For example, an individual could start a rumor cascade by tweeting a story or claim with an assertion in it, and another individual could independently start a second cascade of the same rumor (pertaining to the same story or claim) that is completely independent of the first cascade, except that it pertains to the same story or claim. If they remain independent, they represent two cascades of the same rumor. Cascades can be as small as size one (meaning no one retweeted the original tweet). The number of cascades that make up a rumor is equal to the number of times the story or claim was independently tweeted by a user (not retweeted). So, if a rumor “A” is tweeted by 10 people separately, but not retweeted, it would have 10 cascades, each of size one. Conversely, if a second rumor “B” is independently tweeted by two people and each of those two tweets is retweeted 100 times, the rumor would consist of two cascades, each of size 100.

Here we investigate the differential diffusion of true, false, and mixed (partially true, partially false) news stories using a comprehensive data set of all of the fact-checked rumor cascades that spread on Twitter from its inception in 2006 to 2017. The data include ~126,000 rumor cascades spread by ~3 million people more than 4.5 million times. We sampled all rumor cascades investigated by six independent fact-checking organizations

<sup>1</sup>Massachusetts Institute of Technology (MIT), the Media Lab, E14-526, 75 Amherst Street, Cambridge, MA 02142, USA. <sup>2</sup>MIT, E62-364, 100 Main Street, Cambridge, MA 02142, USA.

\*Corresponding author. Email: sinan@mit.edu

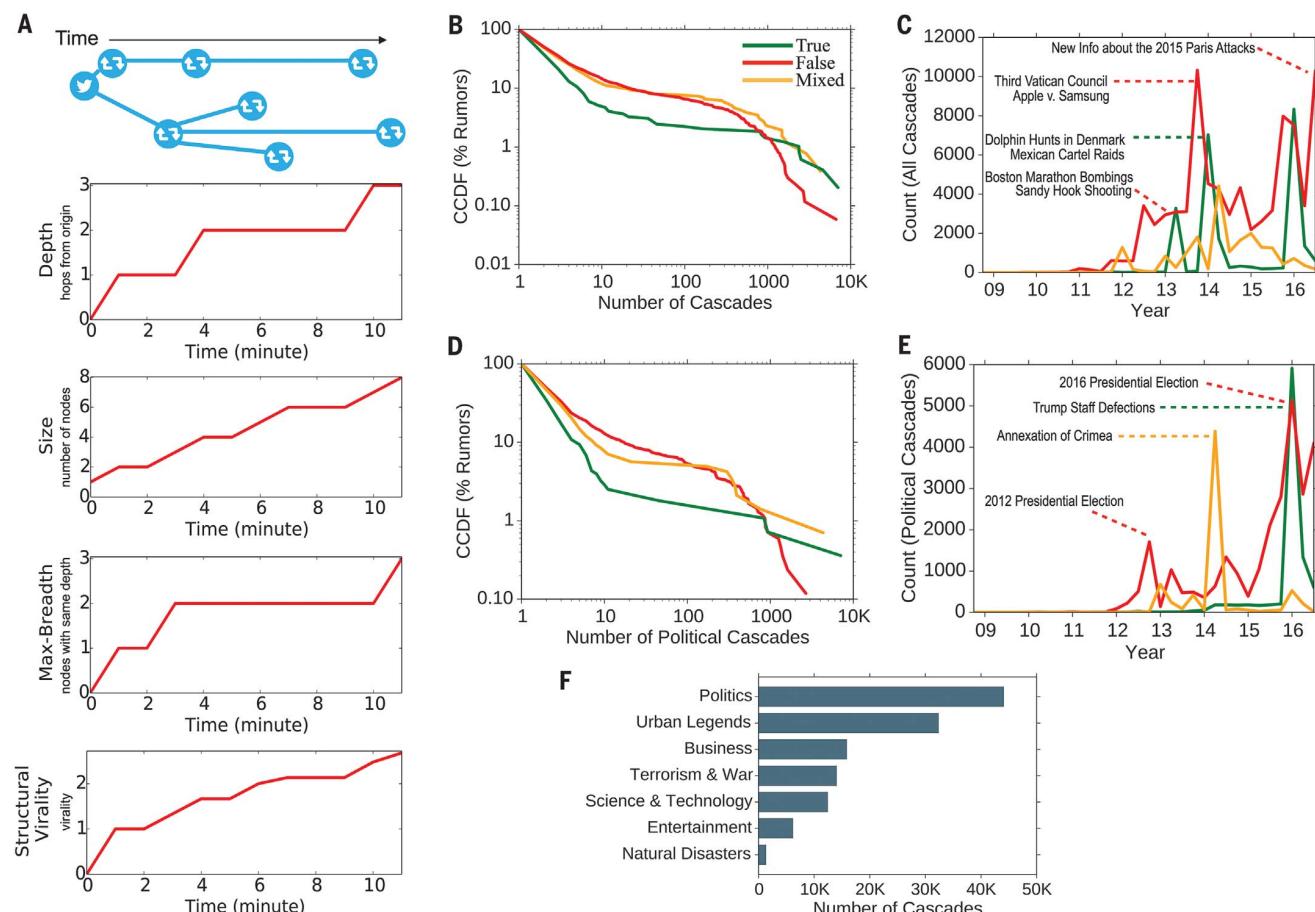
(snopes.com, politifact.com, factcheck.org, truthfiction.com, hoax-slayer.com, and urbanlegends.about.com) by parsing the title, body, and verdict (true, false, or mixed) of each rumor investigation reported on their websites and automatically collecting the cascades corresponding to those rumors on Twitter. The result was a sample of rumor cascades whose veracity had been agreed on by these organizations between 95 and 98% of the time. We cataloged the diffusion of the rumor cascades by collecting all English-language replies to tweets that contained a link to any of the aforementioned websites from 2006 to 2017 and used optical character recognition to extract text from images where needed. For each reply tweet, we extracted the original tweet being replied to and all the retweets of the original tweet. Each retweet cascade represents a rumor propagating on Twitter that has been verified as true or false by the fact-checking organizations (see the supplementary materials for more details on cascade construction). We then quantified the cascades'

depth (the number of retweet hops from the origin tweet over time, where a hop is a retweet by a new unique user), size (the number of users involved in the cascade over time), maximum breadth (the maximum number of users involved in the cascade at any depth), and structural virality (23) (a measure that interpolates between content spread through a single, large broadcast and that which spreads through multiple generations, with any one individual directly responsible for only a fraction of the total spread) (see the supplementary materials for more detail on the measurement of rumor diffusion).

As a rumor is retweeted, the depth, size, maximum breadth, and structural virality of the cascade increase (Fig. 1A). A greater fraction of false rumors experienced between 1 and 1000 cascades, whereas a greater fraction of true rumors experienced more than 1000 cascades (Fig. 1B); this was also true for rumors based on political news (Fig. 1D). The total number of false rumors peaked at the end of both 2013 and 2015 and again at the

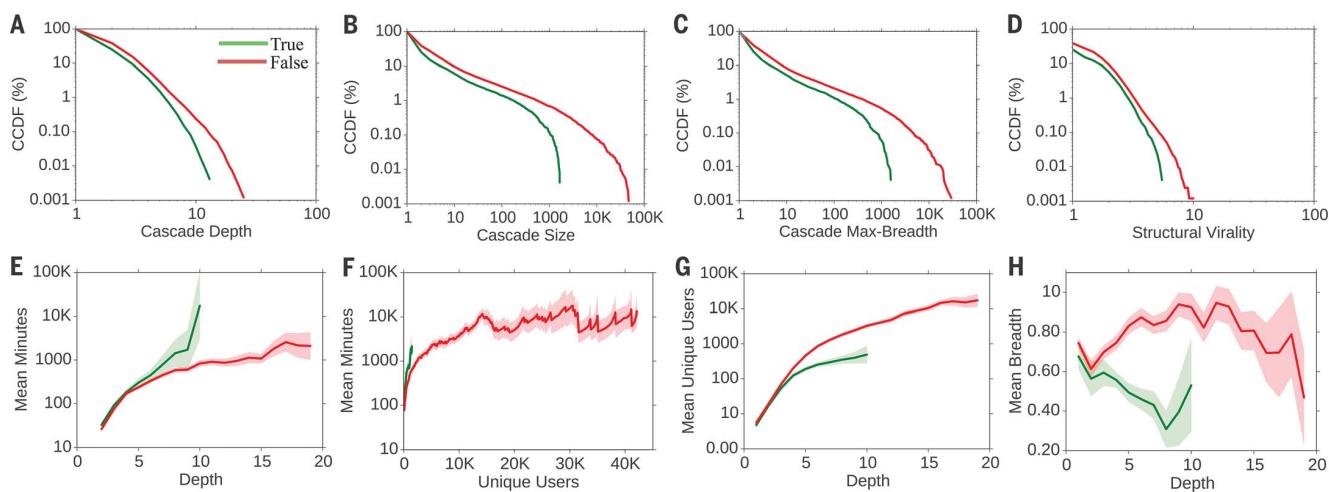
end of 2016, corresponding to the last U.S. presidential election (Fig. 1E). The data also show clear increases in the total number of false political rumors during the 2012 and 2016 U.S. presidential elections (Fig. 1E) and a spike in rumors that contained partially true and partially false information during the Russian annexation of Crimea in 2014 (Fig. 1E). Politics was the largest rumor category in our data, with ~45,000 cascades, followed by urban legends, business, terrorism, science, entertainment, and natural disasters (Fig. 1F).

When we analyzed the diffusion dynamics of true and false rumors, we found that falsehood diffused significantly farther, faster, deeper, and more broadly than the truth in all categories of information [Kolmogorov-Smirnov (K-S) tests are reported in tables S3 to S10]. A significantly greater fraction of false cascades than true cascades exceeded a depth of 10, and the top 0.01% of false cascades diffused eight hops deeper into the Twittersphere than the truth, diffusing to depths



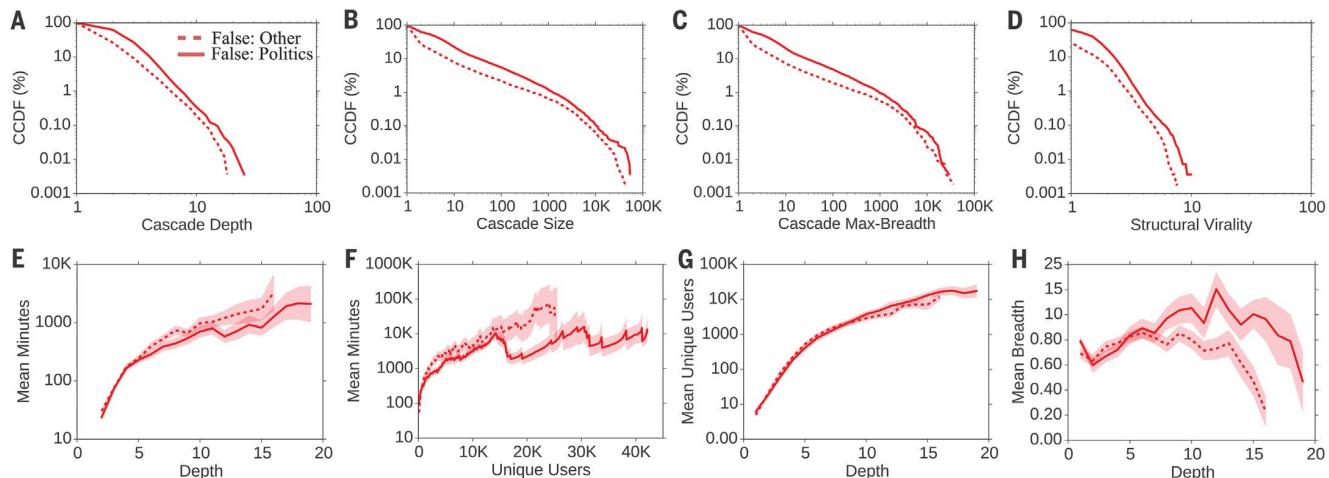
**Fig. 1. Rumor cascades.** (A) An example rumor cascade collected by our method as well as its depth, size, maximum breadth, and structural virality over time. “Nodes” are users. (B) The complementary cumulative distribution functions (CCDFs) of true, false, and mixed (partially true and partially false) cascades, measuring the fraction of rumors that exhibit a given number of cascades. (C) Quarterly counts of all true, false, and mixed rumor cascades

that diffused on Twitter between 2006 and 2017, annotated with example rumors in each category. (D) The CCDFs of true, false, and mixed political cascades. (E) Quarterly counts of all true, false, and mixed political rumor cascades that diffused on Twitter between 2006 and 2017, annotated with example rumors in each category. (F) A histogram of the total number of rumor cascades in our data across the seven most frequent topical categories.



**Fig. 2. Complementary cumulative distribution functions (CCDFs) of true and false rumor cascades.** (A) Depth. (B) Size. (C) Maximum breadth. (D) Structural virility. (E and F) The number of minutes it takes for true and false rumor cascades to reach any (E) depth and (F) number of unique Twitter users. (G) The number of unique Twitter

users reached at every depth and (H) the mean breadth of true and false rumor cascades at every depth. In (H), plot is lognormal. Standard errors were clustered at the rumor level (i.e., cascades belonging to the same rumor were clustered together; see supplementary materials for additional details).



**Fig. 3. Complementary cumulative distribution functions (CCDFs) of false political and other types of rumor cascades.** (A) Depth. (B) Size. (C) Maximum breadth. (D) Structural virility. (E and F) The number of minutes it takes for false political and other false news cascades to reach

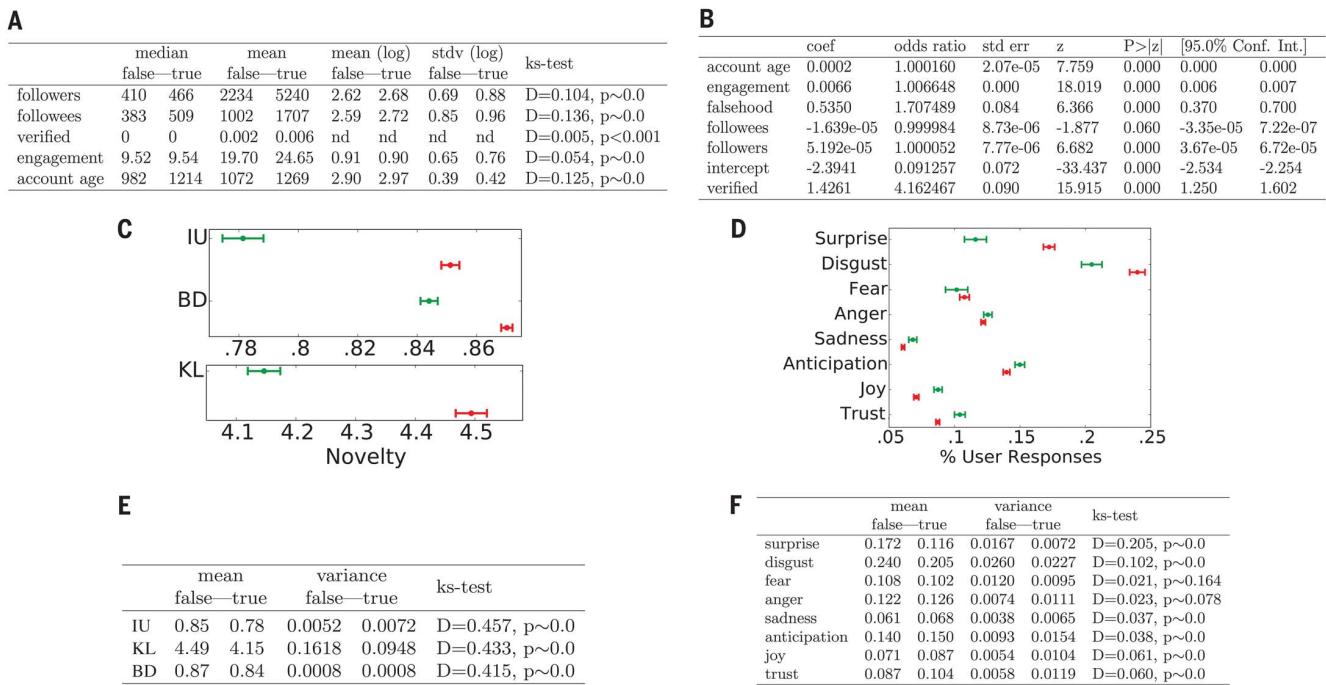
any (E) depth and (F) number of unique Twitter users. (G) The number of unique Twitter users reached at every depth and (H) the mean breadth of these false rumor cascades at every depth. In (H), plot is lognormal. Standard errors were clustered at the rumor level.

greater than 19 hops from the origin tweet (Fig. 2A). Falsehood also reached far more people than the truth. Whereas the truth rarely diffused to more than 1000 people, the top 1% of false-news cascades routinely diffused to between 1000 and 100,000 people (Fig. 2B). Falsehood reached more people at every depth of a cascade than the truth, meaning that many more people retweeted falsehood than they did the truth (Fig. 2C). The spread of falsehood was aided by its virality, meaning that falsehood did not simply spread through broadcast dynamics but rather through peer-to-peer diffusion characterized by a viral branching process (Fig. 2D).

It took the truth about six times as long as falsehood to reach 1500 people (Fig. 2F) and 20 times as long as falsehood to reach a cascade depth of 10 (Fig. 2E). As the truth never diffused beyond a depth of 10, we saw that falsehood reached a depth of 19 nearly 10 times faster than the truth reached a depth of 10 (Fig. 2E). Falsehood also diffused significantly more broadly (Fig. 2H) and was retweeted by more unique users than the truth at every cascade depth (Fig. 2G).

False political news (Fig. 3D) traveled deeper (Fig. 3A) and more broadly (Fig. 3C), reached more people (Fig. 3B), and was more viral than any other category of false information (Fig. 3D). False po-

litical news also diffused deeper more quickly (Fig. 3E) and reached more than 20,000 people nearly three times faster than all other types of false news reached 10,000 people (Fig. 3F). Although the other categories of false news reached about the same number of unique users at depths between 1 and 10, false political news routinely reached the most unique users at depths greater than 10 (Fig. 3G). Although all other categories of false news traveled slightly more broadly at shallower depths, false political news traveled more broadly at greater depths, indicating that more-popular false political news items exhibited broader and more-accelerated diffusion dynamics



**Fig. 4. Models estimating correlates of news diffusion, the novelty of true and false news, and the emotional content of replies to news.**

(A) Descriptive statistics on users who participated in true and false rumor cascades as well as K-S tests of the differences in the distributions of these measures across true and false rumor cascades. (B) Results of a logistic regression model estimating users' likelihood of retweeting a rumor as a function of variables shown at the left. coeff, logit coefficient; z, z score. (C) Differences in the information uniqueness (IU), scaled Bhattacharyya distance (BD), and K-L divergence (KL) of true (green) and false (red) rumor tweets compared to the corpus of prior tweets the user was exposed to in the 60 days before retweeting the rumor tweet. (D) The emotional

content of replies to true (green) and false (red) rumor tweets across seven dimensions categorized by the NRC. (E) Mean and variance of the IU, KL, and BD of true and false rumor tweets compared to the corpus of prior tweets the user has seen in the 60 days before seeing the rumor tweet as well as K-S tests of their differences across true and false rumors. (F) Mean and variance of the emotional content of replies to true and false rumor tweets across seven dimensions categorized by the NRC as well as K-S tests of their differences across true and false rumors. All standard errors are clustered at the rumor level, and all models are estimated with cluster-robust standard errors at the rumor level.

(Fig. 3H). Analysis of all news categories showed that news about politics, urban legends, and science spread to the most people, whereas news about politics and urban legends spread the fastest and were the most viral in terms of their structural virality (see fig. S11 for detailed comparisons across all topics).

One might suspect that structural elements of the network or individual characteristics of the users involved in the cascades explain why falsity travels with greater velocity than the truth. Perhaps those who spread falsity "followed" more people, had more followers, tweeted more often, were more often "verified" users, or had been on Twitter longer. But when we compared users involved in true and false rumor cascades, we found that the opposite was true in every case. Users who spread false news had significantly fewer followers (K-S test = 0.104,  $P \sim 0.0$ ), followed significantly fewer people (K-S test = 0.136,  $P \sim 0.0$ ), were significantly less active on Twitter (K-S test = 0.054,  $P \sim 0.0$ ), were verified significantly less often (K-S test = 0.004,  $P < 0.001$ ), and had been on Twitter for significantly less time (K-S test = 0.125,  $P \sim 0.0$ ) (Fig. 4A). Falsehood

diffused farther and faster than the truth despite these differences, not because of them.

When we estimated a model of the likelihood of retweeting, we found that falsehoods were 70% more likely to be retweeted than the truth (Wald chi-square test,  $P \sim 0.0$ ), even when controlling for the account age, activity level, and number of followers and followees of the original tweeter, as well as whether the original tweeter was a verified user (Fig. 4B). Because user characteristics and network structure could not explain the differential diffusion of truth and falsity, we sought alternative explanations for the differences in their diffusion dynamics.

One alternative explanation emerges from information theory and Bayesian decision theory. Novelty attracts human attention (24), contributes to productive decision-making (25), and encourages information sharing (26) because novelty updates our understanding of the world. When information is novel, it is not only surprising, but also more valuable, both from an information theoretic perspective [in that it provides the greatest aid to decision-making (25)] and from a social perspective [in that it conveys so-

cial status on one that is "in the know" or has access to unique "inside" information (26)]. We therefore tested whether falsity was more novel than the truth and whether Twitter users were more likely to retweet information that was more novel.

To assess novelty, we randomly selected ~5000 users who propagated true and false rumors and extracted a random sample of ~25,000 tweets that they were exposed to in the 60 days prior to their decision to retweet a rumor. We then specified a latent Dirichlet Allocation Topic model (27), with 200 topics and trained on 10 million English-language tweets, to calculate the information distance between the rumor tweets and all the prior tweets that users were exposed to before retweeting the rumor tweets. This generated a probability distribution over the 200 topics for each tweet in our data set. We then measured how novel the information in the true and false rumors was by comparing the topic distributions of the rumor tweets with the topic distributions of the tweets to which users were exposed in the 60 days before their retweet. We found that false rumors were significantly more

novel than the truth across all novelty metrics, displaying significantly higher information uniqueness ( $K\text{-}S$  test = 0.457,  $P \sim 0.0$ ) (28), Kullback-Leibler ( $K\text{-}L$ ) divergence ( $K\text{-}S$  test = 0.433,  $P \sim 0.0$ ) (29), and Bhattacharyya distance ( $K\text{-}S$  test = 0.415,  $P \sim 0.0$ ) (which is similar to the Hellinger distance) (30). The last two metrics measure differences between probability distributions representing the topical content of the incoming tweet and the corpus of previous tweets to which users were exposed.

Although false rumors were measurably more novel than true rumors, users may not have perceived them as such. We therefore assessed users' perceptions of the information contained in true and false rumors by comparing the emotional content of replies to true and false rumors. We categorized the emotion in the replies by using the leading lexicon curated by the National Research Council Canada (NRC), which provides a comprehensive list of ~140,000 English words and their associations with eight emotions based on Plutchik's (31) work on basic emotion—anger, fear, anticipation, trust, surprise, sadness, joy, and disgust (32)—and a list of ~32,000 Twitter hashtags and their weighted associations with the same emotions (33). We removed stop words and URLs from the reply tweets and calculated the fraction of words in the tweets that related to each of the eight emotions, creating a vector of emotion weights for each reply that summed to one across the emotions. We found that false rumors inspired replies expressing greater surprise ( $K\text{-}S$  test = 0.205,  $P \sim 0.0$ ), corroborating the novelty hypothesis, and greater disgust ( $K\text{-}S$  test = 0.102,  $P \sim 0.0$ ), whereas the truth inspired replies that expressed greater sadness ( $K\text{-}S$  test = 0.037,  $P \sim 0.0$ ), anticipation ( $K\text{-}S$  test = 0.038,  $P \sim 0.0$ ), joy ( $K\text{-}S$  test = 0.061,  $P \sim 0.0$ ), and trust ( $K\text{-}S$  test = 0.060,  $P \sim 0.0$ ) (Fig. 4, D and F). The emotions expressed in reply to falsehoods may illuminate additional factors, beyond novelty, that inspire people to share false news. Although we cannot claim that novelty causes retweets or that novelty is the only reason why false news is retweeted more often, we do find that false news is more novel and that novel information is more likely to be retweeted.

Numerous diagnostic statistics and manipulation checks validated our results and confirmed their robustness. First, as there were multiple cascades for every true and false rumor, the variance of error terms associated with cascades corresponding to the same rumor will be correlated. We therefore specified cluster-robust standard errors and calculated all variance statistics clustered at the rumor level. We tested the robustness of our findings to this specification by comparing analyses with and without clustered errors and found that, although clustering reduced the precision of our estimates as expected, the directions, magnitudes, and significance of our results did not change, and chi-square ( $P \sim 0.0$ ) and deviance ( $d$ ) goodness-of-fit tests ( $d = 3.4649 \times 10^{-6}$ ,  $P \sim 1.0$ ) indicate that the models are well specified (see supplementary materials for more detail).

Second, a selection bias may arise from the restriction of our sample to tweets fact checked by the six organizations we relied on. Fact checking may select certain types of rumors or draw additional attention to them. To validate the robustness of our analysis to this selection and the generalizability of our results to all true and false rumor cascades, we independently verified a second sample of rumor cascades that were not verified by any fact-checking organization. These rumors were fact checked by three undergraduate students at Massachusetts Institute of Technology (MIT) and Wellesley College. We trained the students to detect and investigate rumors with our automated rumor-detection algorithm running on 3 million English-language tweets from 2016 (34). The undergraduate annotators investigated the veracity of the detected rumors using simple search queries on the web. We asked them to label the rumors as true, false, or mixed on the basis of their research and to discard all rumors previously investigated by one of the fact-checking organizations. The annotators, who worked independently and were not aware of one another, agreed on the veracity of 90% of the 13,240 rumor cascades that they investigated and achieved a Fleiss' kappa of 0.88. When we compared the diffusion dynamics of the true and false rumors that the annotators agreed on, we found results nearly identical to those estimated with our main data set (see fig. S17). False rumors in the robustness data set had greater depth ( $K\text{-}S$  test = 0.139,  $P \sim 0.0$ ), size ( $K\text{-}S$  test = 0.131,  $P \sim 0.0$ ), maximum breadth ( $K\text{-}S$  test = 0.139,  $P \sim 0.0$ ), structural virality ( $K\text{-}S$  test = 0.066,  $P \sim 0.0$ ), and speed (fig. S17) and a greater number of unique users at each depth (fig. S17). When we broadened the analysis to include majority-rule labeling, rather than unanimity, we again found the same results (see supplementary materials for results using majority-rule labeling).

Third, although the differential diffusion of truth and falsity is interesting with or without robot, or bot, activity, one may worry that our conclusions about human judgment may be biased by the presence of bots in our analysis. We therefore used a sophisticated bot-detection algorithm (35) to identify and remove all bots before running the analysis. When we added bot traffic back into the analysis, we found that none of our main conclusions changed—false news still spread farther, faster, deeper, and more broadly than the truth in all categories of information. The results remained the same when we removed all tweet cascades started by bots, including human retweets of original bot tweets (see supplementary materials, section S8.3) and when we used a second, independent bot-detection algorithm (see supplementary materials, section S8.3.5) and varied the algorithm's sensitivity threshold to verify the robustness of our analysis (see supplementary materials, section S8.3.4). Although the inclusion of bots, as measured by the two state-of-the-art bot-detection algorithms we used in our analysis, accelerated the spread of both true and false news, it affected their spread roughly equally. This suggests that false

news spreads farther, faster, deeper, and more broadly than the truth because humans, not robots, are more likely to spread it.

Finally, more research on the behavioral explanations of differences in the diffusion of true and false news is clearly warranted. In particular, more robust identification of the factors of human judgment that drive the spread of true and false news online requires more direct interaction with users through interviews, surveys, lab experiments, and even neuroimaging. We encourage these and other approaches to the investigation of the factors of human judgment that drive the spread of true and false news in future work.

False news can drive the misallocation of resources during terror attacks and natural disasters, the misalignment of business investments, and misinformed elections. Unfortunately, although the amount of false news online is clearly increasing (Fig. 1, C and E), the scientific understanding of how and why false news spreads is currently based on ad hoc rather than large-scale systematic analyses. Our analysis of all the verified true and false rumors that spread on Twitter confirms that false news spreads more pervasively than the truth online. It also overturns conventional wisdom about how false news spreads. Though one might expect network structure and individual characteristics of spreaders to favor and promote false news, the opposite is true. The greater likelihood of people to retweet falsity more than the truth is what drives the spread of false news, despite network and individual factors that favor the truth. Furthermore, although recent testimony before congressional committees on misinformation in the United States has focused on the role of bots in spreading false news (36), we conclude that human behavior contributes more to the differential spread of falsity and truth than automated robots do. This implies that misinformation-containment policies should also emphasize behavioral interventions, like labeling and incentives to dissuade the spread of misinformation, rather than focusing exclusively on curtailing bots. Understanding how false news spreads is the first step toward containing it. We hope our work inspires more large-scale research into the causes and consequences of the spread of false news as well as its potential cures.

## REFERENCES AND NOTES

1. L. J. Savage, *J. Am. Stat. Assoc.* **46**, 55–67 (1951).
2. H. A. Simon, *The New Science of Management Decision* (Harper & Brothers Publishers, New York, 1960).
3. R. Wedgwood, *Noûs* **36**, 267–297 (2002).
4. E. Fehr, U. Fischbacher, *Nature* **425**, 785–791 (2003).
5. C. E. Shannon, *Bell Syst. Tech. J.* **27**, 379–423 (1948).
6. S. Bikhchandani, D. Hirshleifer, I. Welch, *J. Polit. Econ.* **100**, 992–1026 (1992).
7. K. Rapoza, "Can 'fake news' impact the stock market?" *Forbes*, 26 February 2017; www.forbes.com/sites/kenrapoza/2017/02/26/can-fake-news-impact-the-stock-market/.
8. M. Mendoza, B. Poblete, C. Castillo, in *Proceedings of the First Workshop on Social Media Analytics* (Association for Computing Machinery, ACM, 2010), pp. 71–79.
9. A. Gupta, H. Lamba, P. Kumaraguru, A. Joshi, in *Proceedings of the 22nd International Conference on World Wide Web* (ACM, 2010), pp. 729–736.

10. K. Starbird, J. Maddock, M. Orand, P. Achterman, R. M. Mason, in *iConference 2014 Proceedings* (iSchools, 2014).
11. J. Gottfried, E. Shearer, "News use across social media platforms," Pew Research Center, 26 May 2016; [www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016/](http://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016/).
12. C. Silverman, "This analysis shows how viral fake election news stories outperformed real news on Facebook," *BuzzFeed News*, 16 November 2016; [www.buzzfeed.com/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook/](http://www.buzzfeed.com/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook/).
13. M. De Domenico, A. Lima, P. Mougel, M. Musolesi, *Sci. Rep.* **3**, 2980 (2013).
14. O. Oh, K. H. Kwon, H. R. Rao, in *Proceedings of the International Conference on Information Systems* (International Conference on Information Systems, ICIIS, paper 231, 2010).
15. M. Tambusco, G. Ruffo, A. Flammini, F. Menczer, in *Proceedings of the 24th International Conference on World Wide Web* (ACM, 2015), pp. 977–982.
16. Z. Zhao, P. Resnick, Q. Mei, in *Proceedings of the 24th International Conference on World Wide Web* (ACM, 2015), pp. 1395–1405.
17. M. Gupta, P. Zhao, J. Han, in *Proceedings of the 2012 Society for Industrial and Applied Mathematics International Conference on Data Mining* (Society for Industrial and Applied Mathematics, SIAM, 2012), pp. 153–164.
18. G. L. Ciampaglia et al., *PLOS ONE* **10**, e0128193 (2015).
19. A. Frigeri, L. A. Adamic, D. Eckles, J. Cheng, in *Proceedings of the International Conference on Weblogs and Social Media* (Association for the Advancement of Artificial Intelligence, AAAI, 2014).
20. M. Del Vicario et al., *Proc. Natl. Acad. Sci. U.S.A.* **113**, 554–559 (2016).
21. A. Bessi et al., *PLOS ONE* **10**, e0118093 (2015).
22. Frigeri et al. (19) do evaluate two metrics of diffusion: depth, which shows little difference between true and false rumors, and shares per rumor, which is higher for true rumors than it is for false rumors. Although these results are important, they are not definitive owing to the smaller sample size of the study; the early timing of the sample, which misses the rise of false news after 2013; and the fact that more shares per rumor do not necessarily equate to deeper, broader, or more rapid diffusion.
23. S. Goel, A. Anderson, J. Hofman, D. J. Watts, *Manage. Sci.* **62**, 180–196 (2015).
24. L. Itti, P. Baldi, *Vision Res.* **49**, 1295–1306 (2009).
25. S. Aral, M. Van Alstyne, *Am. J. Sociol.* **117**, 90–171 (2011).
26. J. Berger, K. L. Milkman, *J. Mark. Res.* **49**, 192–205 (2012).
27. D. M. Blei, A. Y. Ng, M. I. Jordan, *J. Mach. Learn. Res.* **3**, 993–1022 (2003).
28. S. Aral, P. Dhillon, "Unpacking novelty: The anatomy of vision advantages," Working paper, MIT-Sloan School of Management, Cambridge, MA, 22 June 2016; [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2388254](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2388254).
29. T. M. Cover, J. A. Thomas, *Elements of Information Theory* (Wiley, 2012).
30. T. Kailath, *IEEE Trans. Commun. Technol.* **15**, 52–60 (1967).
31. R. Plutchik, *Am. Sci.* **89**, 344–350 (2001).
32. S. M. Mohammad, P. D. Turney, *Comput. Intell.* **29**, 436–465 (2013).
33. S. M. Mohammad, S. Kiritchenko, *Comput. Intell.* **31**, 301–326 (2015).
34. S. Vosoughi, D. Roy, in *Proceedings of the 10th International AAAI Conference on Weblogs and Social Media* (AAAI, 2016), pp. 707–710.
35. C. A. Davis, O. Varol, E. Ferrara, A. Flammini, F. Menczer, in *Proceedings of the 25th International Conference Companion on World Wide Web* (ACM, 2016), pp. 273–274.
36. For example, this is an argument made in recent testimony by Clint Watts—Robert A. Fox Fellow at the Foreign Policy

Research Institute and Senior Fellow at the Center for Cyber and Homeland Security at George Washington University—given during the U.S. Senate Select Committee on Intelligence hearing on "Disinformation: A Primer in Russian Active Measures and Influence Campaigns" on 30 March 2017; [www.intelligence.senate.gov/sites/default/files/documents/os-cwatts-033017.pdf](http://www.intelligence.senate.gov/sites/default/files/documents/os-cwatts-033017.pdf).

#### ACKNOWLEDGMENTS

We are indebted to Twitter for providing funding and access to the data. We are also grateful to members of the MIT research community for invaluable discussions. The research was approved by the MIT institutional review board. The analysis code is freely available at <https://goo.gl/forms/AK1lZujpepxhNTy33>. The entire data set is also available, from the same link, upon signing an access agreement stating that (i) you shall only use the data set for the purpose of validating the results of the MIT study and for no other purpose; (ii) you shall not attempt to identify, reidentify, or otherwise deanonymize the data set; and (iii) you shall not further share, distribute, publish, or otherwise disseminate the data set. Those who wish to use the data for any other purposes can contact and make a separate agreement with Twitter.

#### SUPPLEMENTARY MATERIALS

[www.sciencemag.org/content/359/6380/1146/suppl/DC1](http://www.sciencemag.org/content/359/6380/1146/suppl/DC1)  
Materials and Methods  
Figs. S1 to S20  
Tables S1 to S39  
References (37–75)

14 September 2017; accepted 19 January 2018  
10.1126/science.aap9559

## 外 文 译 文

## 真假新闻的在线传播

Soroush Vosoughi, Deb Roy, Sinan Aral

麻省理工学院

# 第一章 概述

## 1.1 概述

决策、合作、通信和市场领域的基础理论全都将对真实或准确度的概念化作为几乎一切人类努力的核心。然而，不论是真实信息还是虚假信息都会于在线媒体上迅速传播。定义什么是真、什么是假成了一种常见的政治策略，而不是基于一些各方同意的事实的争论。我们的经济也难免遭受虚假信息传播的影响。虚假流言会影响股价和大规模投资的动向，例如，在一条声称巴拉克·奥巴马在爆炸中受伤的推文发布后，股市市值蒸发了 1300 亿美元。的确，从自然灾害到恐怖袭击，我们对一切事情的反应都受到了扰乱。

新的社交网络技术在使信息的传播速度变快和规模变大的同时，也便利了不实信息（即不准确或有误导性的信息）的传播。然而，尽管我们对信息和新闻的获取越来越多地收到这些新技术的引导，但我们仍然对他们在虚假信息传播上的作用知之甚少。尽管媒体对假新闻传播的轶事分析给予了相当多的关注，但仍然几乎没有针对不实信息扩散或其发布源头的大规模实证调查。目前，虚假信息传播的研究仅仅局限于小的、局部的样本的分析上，而这些分析忽略了两个最重要的科学问题：真实信息和虚假信息的传播有什么不同？哪些人类判断中的因素可以解释这些不同？

$$\max_W \text{tr}(\mathbf{W}^\top \mathbf{X} \mathbf{X}^\top \mathbf{W}) \quad \text{式 (外 1-1)}$$

的字。我只是为了把第二章挤到下一页而凑的字。我只是为了把第二章挤到下一页而凑的字。

## 第二章 我也不知道是什么

新的社交网络技术在使信息的传播速度变快和规模变大的同时，也便利了不实信息（即不准确或有误导性的信息）的传播。然而，尽管我们对信息和新闻的获取越来越多地收到这些新技术的引导，但我们仍然对他们在虚假信息传播上的作用知之甚少。尽管媒体对假新闻传播的轶事分析给予了相当多的关注，但仍然几乎没有针对不实信息扩散或其发布源头的大规模实证调查。目前，虚假信息传播的研究仅仅局限于小的、局部的样本的分析上，而这些分析忽略了两个最重要的科学问题：真实信息和虚假信息的传播有什么不同？哪些人类判断中的因素可以解释这些不同？

新的社交网络技术在使信息的传播速度变快和规模变大的同时，也便利了不实信息（即不准确或有误导性的信息）的传播。然而，尽管我们对信息和新闻的获取越来越多地收到这些新技术的引导，但我们仍然对他们在虚假信息传播上的作用知之甚少。尽管媒体对假新闻传播的轶事分析给予了相当多的关注，但仍然几乎没有针对不实信息扩散或其发布源头的大规模实证调查。目前，虚假信息传播的研究仅仅局限于小的、局部的样本的分析上，而这些分析忽略了两个最重要的科学问题：真实信息和虚假信息的传播有什么不同？哪些人类判断中的因素可以解释这些不同？

新的社交网络技术在使信息的传播速度变快和规模变大的同时，也便利了不实信息（即不准确或有误导性的信息）的传播。然而，尽管我们对信息和新闻的获取越来越多地收到这些新技术的引导，但我们仍然对他们在虚假信息传播上的作用知之甚少。尽管媒体对假新闻传播的轶事分析给予了相当多的关注，但仍然几乎没有针对不实信息扩散或其发布源头的大规模实证调查。目前，虚假信息传播的研究仅仅局限于小的、局部的样本的分析上，而这些分析忽略了两个最重要的科学问题：真实信息和虚假信息的传播有什么不同？哪些人类判断中的因素可以解释这些不同？

$$\max_{\mathbf{W}} \text{tr}(\mathbf{W}^T \mathbf{X} \mathbf{X}^T \mathbf{W}) \quad \text{式 (外 2-1)}$$

新的社交网络技术在使信息的传播速度变快和规模变大的同时，也便利了不实信息（即不准确或有误导性的信息）的传播。然而，尽管我们对信息和新闻的获取越来越多地收到这些新技术的引导，但我们仍然对他们在虚假信息传播上的作用知之甚少。尽管媒体对假新闻传播的轶事分析给予了相当多的关注，但仍然几乎没有针对不实信息扩散或其发布源头的大规模实证调查。目前，虚假信息传播的研究仅仅局限于小的、局部的样本的分析上，而这些分析忽略了两个最重要的科学问题：真实信息和虚假信息的传播有什么不同？哪些人类判断中的因素可以解释这些不同？

新的社交网络技术在使信息的传播速度变快和规模变大的同时，也便利了不实信息（即不准确或有误导性的信息）的传播。然而，尽管我们对信息和新闻的获取越来越多地收到这些新技术的引导，但我们仍然对他们在虚假信息传播上的作用知之甚少。尽管媒体对假新闻传播的轶事分析给予了相当多的关注，但仍然几乎没有针对不实信息扩散或其发布源头的大规模实证调查。目前，虚假信息传播的研究仅仅局限于小的、局部的

样本的分析上，而这些分析忽略了两个最重要的科学问题：真实信息和虚假信息的传播有什么不同？哪些人类判断中的因素可以解释这些不同？

新的社交网络技术在使信息的传播速度变快和规模变大的同时，也便利了不实信息（即不准确或有误导性的信息）的传播。然而，尽管我们对信息和新闻的获取越来越多地收到这些新技术的引导，但我们仍然对他们在虚假信息传播上的作用知之甚少。尽管媒体对假新闻传播的轶事分析给予了相当多的关注，但仍然几乎没有针对不实信息扩散或其发布源头的大规模实证调查。目前，虚假信息传播的研究仅仅局限于小的、局部的样本的分析上，而这些分析忽略了两个最重要的科学问题：真实信息和虚假信息的传播有什么不同？哪些人类判断中的因素可以解释这些不同？

# 北京邮电大学

## 本科毕业设计（论文）开题报告

学院	信息与通信工程学院	专业	通信工程	班级	201421119
学生姓名	猜猜	学号	2014210999	班内序号	99
指导教师姓名	猜猜	所在单位	信息与通信工程学院	职称	教授
设计（论文） 题目	(中文) 猜猜看毕设题目是什么 (英文) Just Guess What On Earth My Title is				

### 一、选题背景及意义

社交多媒体(social multimedia)是多媒体数据(multimedia)与社交媒体(social media)相结合的新型媒体形式。它是互联网技术发展过程中，人们对多样的媒体内容和新型的交互模式的需求中产生的。其中，多媒体数据极大地丰富了纯文本内容，而社会媒体网络提供了快速交流、传播多媒体内容的高效平台，两者相互转化。全世界内，最引人注目的社交媒体平台当属微博客(Microblog)，其中以中文的新浪微博和英文的Twitter最为活跃，各平台每时每刻产生并流动着种类繁多的大量信息。

微博客平台有着发布方便、传播迅速、受众广泛且总量大的特点。这种特点使得更多的官方媒体将其作为资讯发布的重要平台，同时更多的普通用户将其作为获取热点信息的重要来源。然而，在加速真实信息的有效传播的同时，微博客平台也成了虚假消息的温床，这一现象在社会和科学健康类话题中表现突出：在重大事件、突发事件和灾害事故消息等社会类话题中，虚假信息的传播严重扰乱了网络空间秩序，冲击着网民的认知，有的甚至导致了社会秩序的混乱(如日本福岛核电站泄露事件发生后我国的食用盐哄抢事件)和事件走向的转变(如2016年的美国总统选举)；在科学健康类话题中，耸人听闻的食品安全曝光(如“塑料紫菜”、“棉花肉松”)、不科学的食品安全警告(如“柿子和酸奶一起吃会中毒致死”)和错误的医疗手段(如“一滴血就能验癌”)极易对人们的认知造成误导，进一步带来不必要的麻烦和相应的经济冲击。

### 二、研究的基本内容

对所提出算法进行性能的测试、比较和分析，针对结论面向未来发展方向进行探讨。

### 三、 研究方法及措施

从数据分布的角度上讲，检测谣言的这一类问题非常适合归入数据挖掘的经典问题——异常检测（anomaly detection）或离群点检测（outlier detection），一方面是因为谣言的种类繁多，若归入一大类，其与正常信息的边界可能会难以寻找；另一方面是即便虚假信息被认为泛滥成灾，但谣言在微博空间中仍是少数，可获取的谣言和非谣言比例失衡。

### 四、 研究工作的步骤与进度

2018.1.1 ~ 2018.2.10 完成领域内容调研，模板对应部分撰写。

2018.2.28~2018.4.15 完成相关模板研究，设计模板。

2018.4.16~2018.4.30 进行模板设计评估和比较分析。

2018.5.1~2018.5.15 模板整体撰写。

### 五、 主要参考文献

Zubiaga A, Aker A, Bontcheva K, et al. Detection and Resolution of Rumours in Social Media: A Survey[J]. ACM Computing Surveys (CSUR), 2018, 51(2): 32.

Savage D, Zhang X, Yu X, et al. Anomaly detection in online social networks[J]. Social Networks, 2014, 39(1):62-70.

Castillo C, Mendoza M, Poblete B. Information credibility on twitter[C]// International Conference on World Wide Web, WWW 2011, Hyderabad, India, March 28 - April. DBLP, 2011:675-684.

Jin Z, Cao J, Guo H, et al. Multimodal Fusion with Recurrent Neural Networks for Rumor Detection on Microblogs[C]//Proceedings of the 2017 ACM on Multimedia Conference. ACM, 2017: 795-816.

指导教师签字		日期	年 月 日
--------	--	----	-------

注：可根据开题报告的长度加页。

**北京邮电大学**  
**本科毕业设计（论文）中期进展情况检查表**

学院	信息与通信工程学院					
专业	通信工程					
班级	2014211199					
学生姓名	猜猜					
学号	2014210999					
指导教师姓名						
所在单位						
职称						
设计（论文）题目	(中文) 猜猜看毕设题目是什么					
	(英文) Just Guess What On Earth My Title is					
目前已完成任务	<p>截至中期检查前夕，本课题已经完成的工作如下：</p> <p><b>完成有关实验。</b></p> <p>实验结束后，对整体准确率（Accuracy）进行了统计，还得到了谣言和非谣言的精度（Precision）、召回率（Recall）和F1值（F1-Score）。</p>					
	是否符合任务书要求进度 是					
尚需完成的任务	<ul style="list-style-type: none"> <li>• 完成整体架构和论文书写任务。</li> <li>• 完成外文文献的翻译。</li> </ul>					
	能否按期完成设计（论文） 能					
存在问题和解决办法	存在 问题	模型中存在一些欠缺讨论分析的细节，如阈值选择等。				
	拟采取的办法	拟补充部分实验和查阅领域经验进行讨论分析。				
指导教师签字		日期	年   月   日			

检查小 组意见	
负责人签字:      年 月 日	

注: 可根据长度加页。

北 京 邮 电 大 学

教师指导本科毕业设计（论文）记录表

学院	专业	班级	
学生姓名	学号	班内序号	
指导教师姓名	职称		
第 1—2 周记录:			
指导教师签字		日期	年 月 日
第 3—4 周记录:			
指导教师签字		日期	年 月 日

注：每 2 周指导内容记录在一个表格中，双面打印。