

## 第02讲 学术规范

信息学院 (智能应用研究院)

欧新宇



雲南財經大學

YUNNAN UNIVERSITY OF FINANCE AND ECONOMICS

大纲  
Contents

- 01 为什么写作很重要?
- 02 如何定义学术不端
- 03 什么是剽窃?
- 04 如何避免学术不端
- 05 案例分析

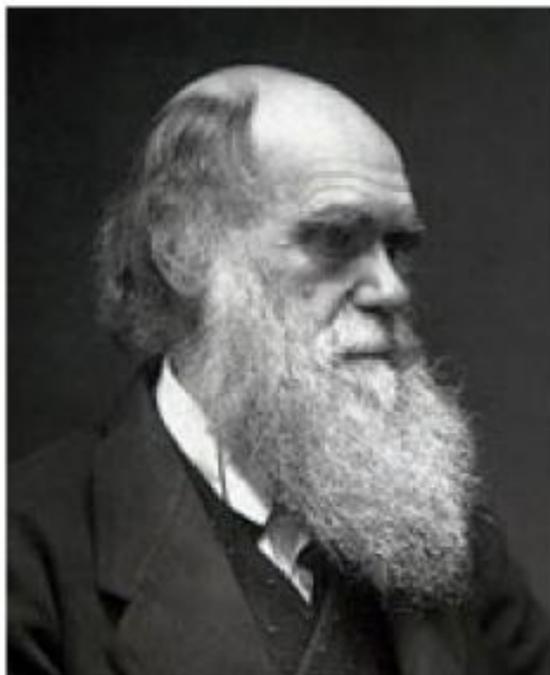
# Part 01

## 为什么写作很重要?

- / 科学的本质
- / 什么是研究
- / 写论文有什么用

# 为什么写作很重要?

## 科学的本质



**Charles Darwin (1809-1882)**

**Science: “... science consists in grouping facts so that general laws or conclusions may be drawn from them.”**

**科学: “科学就是把事实归类, 以便从中得出一般规律或结论。”**

# 为什么写作很重要?

## 什么是研究?

*Research = Re + search*

- **科学研究是创造、重组和更新知识的过程，也是在新领域中创造性地运用知识的过程。**
  - **新颖性 (Novelty) : 创造 > 更新 > 在新领域中的应用**
  - **重组 (Re-reorganizing) : 综述论文**

# 为什么写作很重要?

## 什么是研究?

**Academic paper:** " ... contains **original research results or reviews existing results**. Such a paper ... will only be considered valid if it **undergoes a process of peer review** by one or more referees (who are academics in the same field) who check that the content of the paper is suitable for publication in the journal."

**学术论文:** ".....包含**原创的研究成果或综述现有的研究成果**。这样的论文.....只有在经过一个或多个评审员（同一领域的学者）的**同行评审**过程，确认论文内容适合在该期刊上发表后，才会被视为有效。"

# 为什么写作很重要?

## 写论文有什么用

- 有利于推动同行、社区和业界的发展
- 有利于你个人的职业生涯发展
  - 写作能力对研究工作来说显然很重要
  - 写作能力对其他工作也同样重要
    - ✓ 提出新项目/新观点的能力
    - ✓ 安排研究计划的能力
    - ✓ 证明结果的能力
    - ✓ 一个小项目来模拟你的职业发展

# Part 02

## 如何定义学术不端

*回报、利益较大的问题, 总是存在风险*

/ 国家对学术不端极其重要

/ 学术不端的影响

/ 学术体系的运作机制

/ 学术不端的定义及基本原则

/ 典型学术不端行为

/ 导致学术不端的因素

# 如何定义学术不端

## 国家对于学术不端的问题十分重视

The screenshot shows the official website of the Ministry of Education of the People's Republic of China. The header includes the national emblem and the ministry's name in Chinese and English. A search bar is visible in the top right. The main content area displays the following information:

当前位置: 首页 > 公开

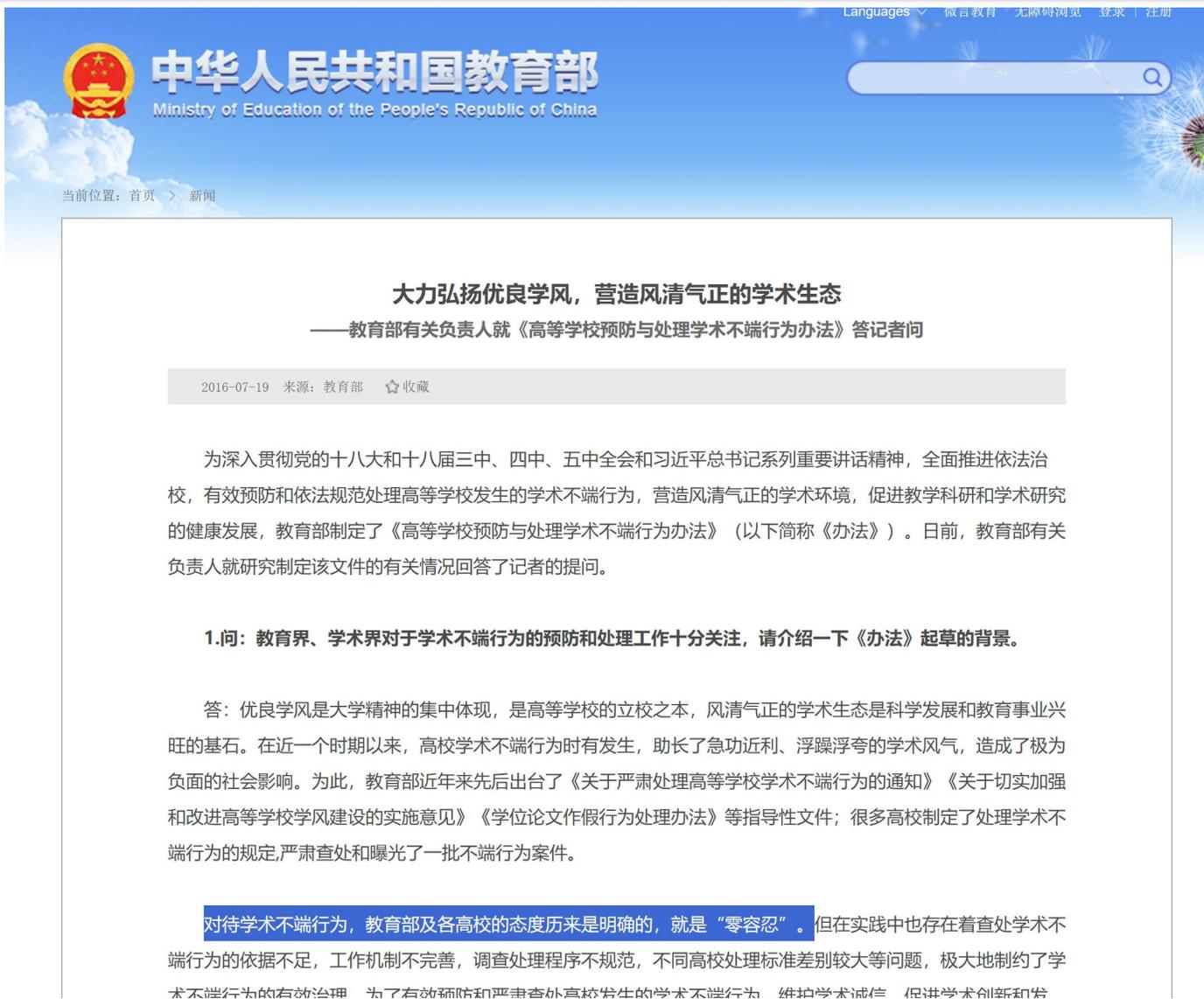
信息名称: 高等学校预防与处理学术不端行为办法  
信息索引: 360A02-03-2016-0018-1 生成日期: 2016-06-16 发文机构: 中华人民共和国教育部  
发文字号: 中华人民共和国教育部令 信息类别: 部门规章 第40号  
内容概述: 《高等学校预防与处理学术不端行为办法》经教育部2016年第14次部长办公会议审议通过, 自2016年9月1日起施行。

**高等学校预防与处理学术不端行为办法**

中华人民共和国教育部令第40号

[http://www.moe.gov.cn/srcsite/A02/s5911/moe\\_621/201607/t20160718\\_272156.html](http://www.moe.gov.cn/srcsite/A02/s5911/moe_621/201607/t20160718_272156.html)

# 如何定义学术不端



教育部：“对待学术不端行为，教育部及各高校的态度历来是明确的，就是**“零容忍”**。

但在实践中也存在着查处学术不端行为的依据不足……”

[http://www.moe.gov.cn/jyb\\_xwfb/s271/201607/t20160719\\_272240.html](http://www.moe.gov.cn/jyb_xwfb/s271/201607/t20160719_272240.html)

# 如何定义学术不端

## 学术不端的影响



学术不端 = 核弹级攻击

# 如何定义学术不端

## 学术体系的运作机制



**学术体系的基础是信任体系，因此对于不诚信的行为严厉到残酷**

## 学术不端的定义

any action or attempted action that may result in **creating an unfair academic advantage** for oneself or an **unfair academic advantage or disadvantage** for any other member or members of the academic community.

任何行为或企图行为，可能导致为自己**创造不公平的学术优势**，或为学术团体的任何其他成员**创造不公平的学术优势或劣势**。

## 学术行为的基本原则

- **尊重**知识产权!
- **明确的引用!**
- **有你自己的贡献!**
- **尊重原创!**
- **不谋取不正当利益**

## 典型学术不端行为

- 抄袭，包括自我抄袭
- 篡改数据
  - 使用**修改或选择**的数据 → 改变解释
  - 篡改他人的学术成果（特别是在新数据集上或不同的参数上）
- 不恰当的作者署名
- 一稿多投

## 不恰当的作者署名

- **排除做出实质性贡献的作者**
- **包括未做出实质性贡献的作者**
- **未经同意就擅自添加作者**
- **作者署名必须得到授权**
- **作者排名与其贡献不符**
- **虚假信息：作者的职位、组织、教育背景、研究经验等**

## 一稿多投

- **提交相同或略有不同的论文**
  - **同时提交给多个期刊**
  - **在撤回或收到决定之前提交给新期刊**
- **期刊延期必须提供**
  - **明确的解释区别**
  - **明确的引用原版 (会议版)**
  - **足够的新内容**

## 导致学术不端行为的因素

- 不好的学习习惯
- 无效的时间管理技能或任务过多
- 心理因素：例如，对自己有不切实际的期望、谎称疾病
- 不知道边界
- 永远不要简单地复制粘贴（例如句子、代码）

# Part 02

## 剽窃

组会是获得反馈、研究方向、建议等最主要的机会。

/ 什么是剽窃

/ 剽窃的原因

/ 学会交流

## 什么是剽窃？

- 剽窃被定义为 “使用他人创作的智力成果而未注明其来源。”
- 为什么这很重要？
  - 只有尊重知识，知识才可能发展
  - 创造知识远比使用知识，甚至是剽窃知识难很多

## 剽窃的原因

- 写作技能信心不足
- 对写作有不切实际的期望
- 笔记质量差、粗心或被动

## 剽窃的原因

- 规范引用不是很容易处理的
- 不要完稿后再引用，应该在编辑时引用，甚至是先标注再引用
- 尽量使用间接引用，少用直接应用，使用重写替代原话引用
- 如果你剽窃，即使只有一次，你也会被开除

## 如何正确引用参考文献

尊重/保护原始来源的信誉。

- 不要断章取义
- 仅限于论证的必要性，不要大篇幅引用
- 尽可能追溯到原始作者
- 引用未发表的作品应获得授权（但建议不要引用）
- 引用口头陈述应提供证据（但建议不要引用）
- 引用应伴有明显的标记（ "" , [1]）
- 如果可能，请引用经过同行评审的期刊论文：**顶会论文** > = 期刊 > 会议 > arXiv

## 如何正确引用参考文献

### 正确的引用

- **缺少相关引用 (假装不知道相关工作)**
  - 有意避免比较
  - 假装创新
- **避免虚假引用 (过多的不恰当引用)**
  - 引用你朋友的不相关论文
  - 引用你不熟悉的文献, 甚至是不存在的文献
  - 引用与你的论文不相关的自我引用
  - 对以前工作的间接引用

## 避免自我剽窃

### 一稿多投

- 在没有适当引用的情况下，为新论文重复使用相似或重复的内容。
- 侵犯版权
- 在期刊扩展中做出适当的引用

## 避免自我剽窃 (A Example)

- 直博生再拿一个硕士学位是有益还是有害?
- 规则的改变, Why?
- 因各种原因, 硕士论文和博士论文产生了重复 -> 自我剽窃
- 学位撤销

**并非所有学术不端行为都是故意的!**

# 剽窃

## 如何识别剽窃?

<b>学生版本是否从别人的原始材料中借用了想法?</b>				
<b>Yes</b>		<b>No</b>		<b>No</b>
<b>至少有一个想法是从原始资料中直接引用了7个或更多的单词</b>				<b>没有抄袭</b>
<b>Yes</b>		<b>No</b>		<b>没有逐字抄袭 和 转述性抄袭</b>
<b>直接逐字引用</b> 是否缺少以下任何内容 <ul style="list-style-type: none"> <li>● 引号</li> <li>● 包含特定定位符的完整文中引用</li> <li>● 参考文献?</li> </ul>		<b>转述的想法</b> 是否缺少以下任何一项 <ul style="list-style-type: none"> <li>● 文中引用</li> <li>● 参考文献?</li> </ul>		
<b>Yes</b>	<b>No</b>	<b>Yes</b>	<b>No</b>	
<b>逐字抄袭</b>	<b>没有逐字抄袭</b>	<b>转述性抄袭</b>	<b>非转述性抄袭 (即直接抄袭)</b>	<b>因此, 这段文字必须是常识或作者自己的想法</b>

Criteria used for Indiana University: <https://plagiarism.iu.edu/IUcriteria.html> (在线测试)

# Part 02

## 如何避免学术不端

*组会是获得反馈、研究方向、建议等最主要的机会。*

/ 学会学习

/ 学会总结

/ 学会交流

# 如何避免学术不端

## 学术不端与刑事犯罪

### 学术不端

- 终身惩罚
- 自证清白

### 严重刑事犯罪

- 严厉惩罚
- 无罪推定

**为你自己提供证据**

# 如何避免学术不端

## 相似的观点总有可能发生 如何为自己提供证据?

### 通过好习惯避免学术不端

- 可信的证据：
  - 带有可信时间戳的正式记录。
  - 例如，私人GitHub/Overleaf项目，以及电子邮件
- 坚持开源
- 可信赖的、安全的合作者

# 如何避免学术不端

## 通过好习惯避免学术不端 (A Example)

示例：(Liu et. al. NPAR' 17) 与 (Liao et. al. GitHub' 17)

- 相似的想法很可能同时出现。
  - Depth-Preserving Style Transfer, Ruizhi Liao, GitHub, 2017
  - Depth-aware Neural Style Transfer, Xiao-Chang Liu et. al., NPAR, 2017.
- 相似性：核心思想、方程、流程图、论文结构、使用 (Chen, et. al., NIPS 2016) 进行深度估计。
- 原因：两项工作都遵循 (Johnson et. al. ECCV' 16)
- 证明：有大量的电子邮件讨论记录、中间结果等。
- 好习惯：使用私人GitHub/Overleaf项目，并保留电子邮件记录。

# 如何避免学术不端

## 通过好习惯避免学术不端 (A Example)

**保持清醒：什么才是真正重要的？**

- 更多的CVPR论文吗？
- 更多的引用次数吗？
- 你解决了什么问题？

**#papers → EI → SCI → ESI → 代表作**

**评估将会更加合理！**

## Copyright 版权问题

- 在引用社交网络数据的时候要小心版权问题



Fig.: Vectorizing Cartoon Animations, IEEE TVCG 2019

考虑一个工具: **Flicker - Creative Commons license**

<https://doresearch.stanford.edu/policies/research-policy-handbook/intellectual-property/copyright-policy>

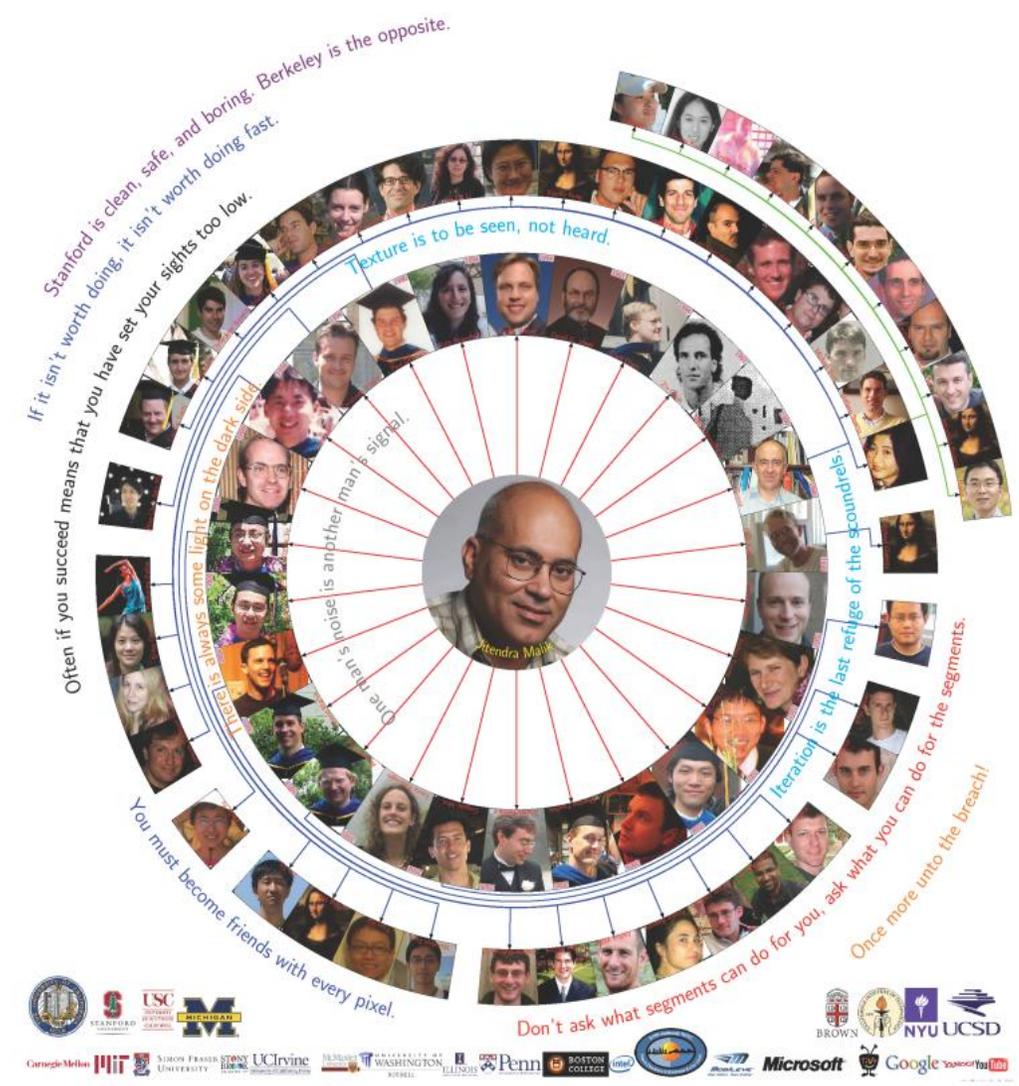


Fig.: J. Malik: "If it is not worth doing, it isn't worth doing fast".

<https://people.eecs.berkeley.edu/~malik/student-tree-2010.pdf>

Part  
02

# 案例分析

九个案例，引以为戒

# 案例分析

## Case1: 皮耶罗·艾佛萨 (Piero Anversa)

### Harvard and the Brigham recommend 31 retractions for cardiac stem cell work

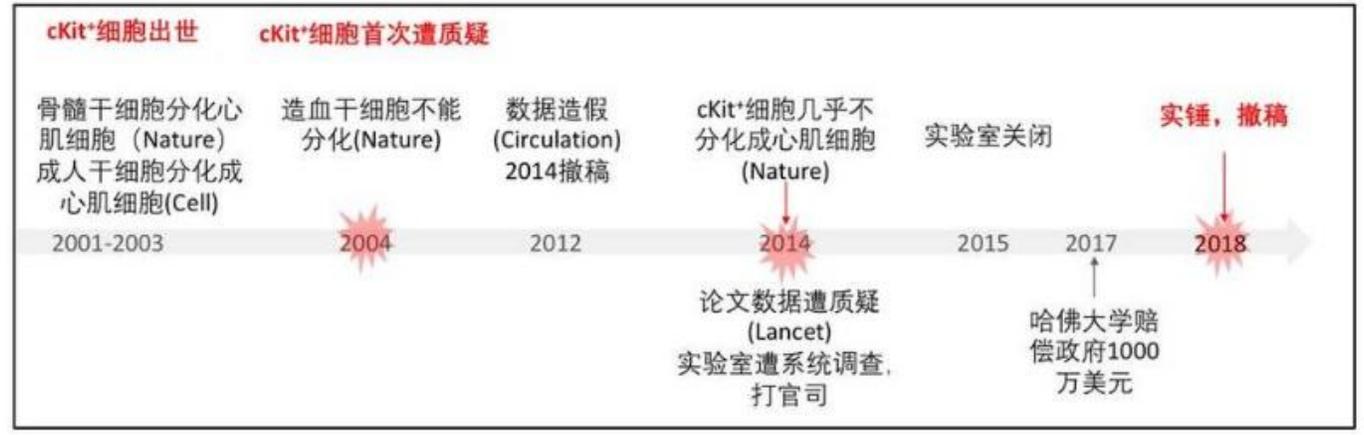
Retraction Watch readers may be familiar with the name Piero Anversa. Until several years ago, Anversa, a scientist at Harvard Medical School and the Brigham and Women's Hospital, was a powerful figure in cardiac stem cell research.



Piero Anversa

“For ten years, he ran everything,” says Jeffery Molkenstin, a researcher at Cincinnati Children's whose lab was among the first to question the basis of Anversa's results in a 2014 paper in Nature.

来源: 撤稿观察网站《Retraction Watch》



[nature](#) > [letters](#) > [article](#)

Letter | Published: 05 April 2001

### Bone marrow cells regenerate infarcted myocardium

[Donald Orlic](#), [Jan Kajstura](#), [Stefano Chimenti](#), [Igor Jakoniuk](#), [Stacie M. Anderson](#), [Baosheng Li](#), [James Pickel](#), [Ronald McKay](#), [Bernardo Nadal-Ginard](#), [David M. Bodine](#), [Annarosa Leri](#) & [Piero Anversa](#) ✉

[Nature](#) **410**, 701–705 (2001) | [Cite this article](#)

19k Accesses | 4082 Citations | 217 Altmetric | [Metrics](#)

## Case2: 小保方晴子

### 事件详情

播报 编辑

#### “成果”发布

2014年1月底,小保方晴子带领的课题组宣布,成功制作出一种全新“万能细胞”STAP。这种细胞定名为“刺激触发采集多功能(Stimulus Triggered Acquisition of Pluripotency)”细胞,简称为STAP细胞。<sup>[3]</sup>

其制作方法很简单,将从老鼠脾脏取出的细胞放在与红茶酸碱度近似的弱酸性溶液里浸泡25分钟,培养数日后就出现了新细胞。此前科学界一直认为,一旦细胞的功能固定下来,在这种程度的刺激下是不可能变成“万能细胞”的。因此,这一成果被认为是颠覆生命科学常识的划时代重大成果。2014年1月29日,这项研究成果刊登在英国科学杂志《自然》电子版上。<sup>[1]</sup>

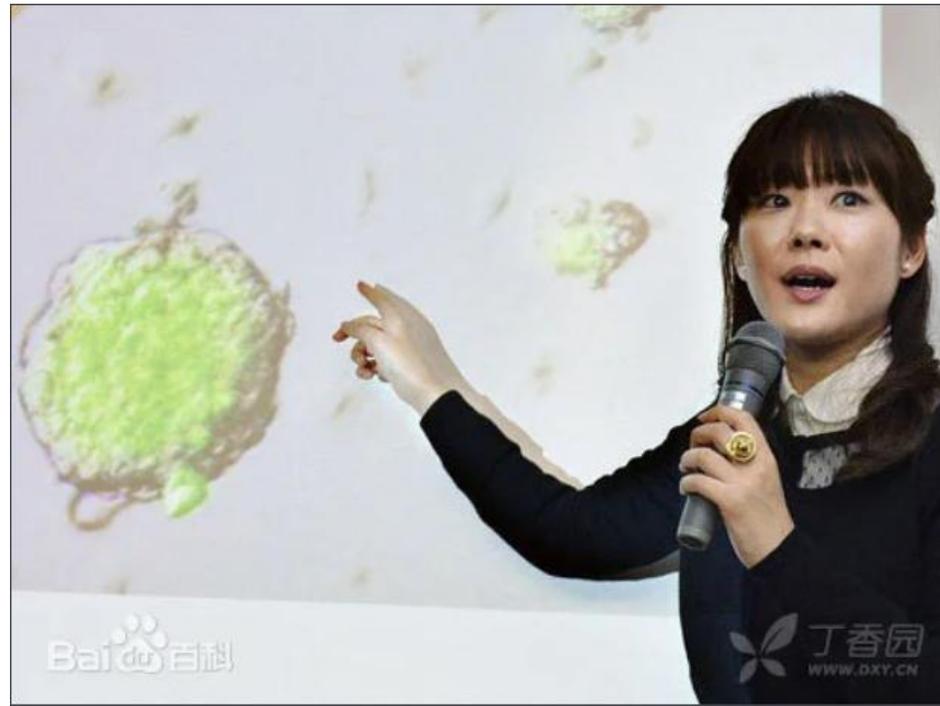
研究小组负责人小保方晴子说:与iPS细胞等技术不同,这项创新技术的亮点是,仅仅通过改变外部环境,给予细胞刺激,就能使细胞发生变化。她认为,这项技术应该能在再生医疗和免疫研究等领域作出贡献。<sup>[4]</sup>

#### 论文被疑造假

小保方晴子论文发表不久后,有外部专家指出,论文中的图像不自然,疑似被加工过。一些国外同行用论文介绍的方法重复实验,却无法再现结果。<sup>[1]</sup>

#### 日媒发布结果确认造假

日本媒体2014年6月3日发表的报道称,在对STAP实验中用到的细胞进行了基因检测后,结果显示,不存在。



STAP细胞事件

## 案例分析

## Case3: 匈牙利总统施密特帕尔

## 施米特·帕尔

匈牙利共和国前总统

播报

编辑

讨论

上传视频



施米特·帕尔（**匈牙利语**：Schmitt Pál，1942年05月13日-），匈牙利政治家，前**击剑**运动员。曾任**匈牙利共和国**总统。

<sup>[1]</sup>1968年、1972年曾两次获得**奥运会男子重剑**团体比赛金牌。35岁退役。先后任匈牙利奥委会主席和国际奥委会副主席以及任匈牙利驻**西班牙**和瑞士大使。2001年竞选国际奥委会主席失利之后转向政坛。2010年5月14日，施密特出任匈牙利国会主席，6月29日当选匈牙利总统。2012年初，被控博士论文涉嫌抄袭，后经查证属实，被剥夺博士学位。施米特因**论文抄袭**事件，于2012年4月2日宣布**辞职**，并获国会批准。

[https://baike.baidu.com/item/%E6%96%BD%E7%B1%B3%E7%89%B9%C2%B7%E5%B8%95%E5%B0%94/6131871?fr=ge\\_al](https://baike.baidu.com/item/%E6%96%BD%E7%B1%B3%E7%89%B9%C2%B7%E5%B8%95%E5%B0%94/6131871?fr=ge_al)

# 案例分析

## Case4: 不只是文字，图片也要避免抄袭

### Android Malware Detection: Multi Model Collaborative Detection System Based on Credibility

1<sup>st</sup> Yufan Li  
College of Cyber Science  
Nankai University

2<sup>nd</sup> Yiwen Liu  
College of Computer science  
Nankai University

3<sup>rd</sup> Guoqing Zhao  
College of Cyber Science  
Nankai University

4<sup>th</sup> Kai Cheng  
College of Cyber Science  
Nankai University

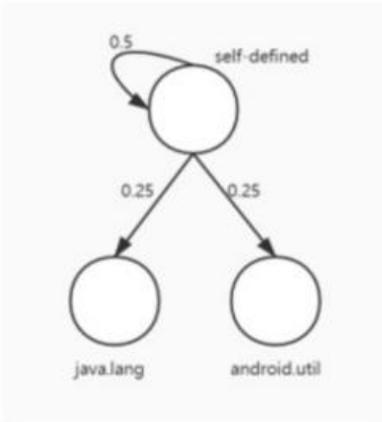


Fig. 1. Markov Chain Example

### MAMADROID: Detecting Android Malware by Building Markov Chains of Behavioral Models\*

Enrico Mariconti<sup>†</sup>, Lucky Onwuzurike<sup>†</sup>, Panagiotis Andriotis<sup>‡</sup>,  
Emiliano De Cristofaro<sup>†</sup>, Gordon Ross<sup>†</sup>, and Gianluca Stringhini<sup>†</sup>  
<sup>†</sup>University College London    <sup>‡</sup>University of the West of England

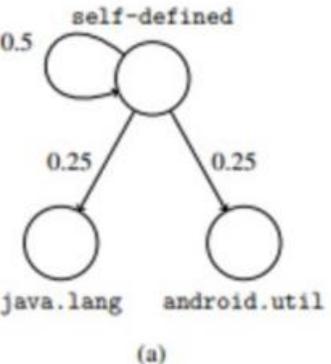


Fig. 5: Markov chains originating from Section II-C when using packages (a)

## 案例分析

## Case5: 同一个作者也有学术不端的风险

## UnrealText: Synthesizing Realistic Scene Text Images from the Unreal World

Shangbang Long  
Carnegie Mellon University  
shangba1@cs.cmu.edu

Cong Yao  
Megvii (Face++) Technology Inc.  
yaocong2010@gmail.com

而另一篇疑似与它相同的论文是**2020年1月15日**发表在《中国科学-信息科学》(Science China) 上的**SynthText3D: synthesizing scene text images from 3D virtual worlds**, 作者: Minghui Liao, Boyu Song, Shangbang Long, Minghang He, Cong Yao & Xiang Bai, 其中**Shangbang Long**、**Cong Yao**与CVPR论文作者相同。

## Fake news!

## SynthText3D: Synthesizing Scene Text Images from 3D Virtual Worlds

<sup>1</sup>Minghui Liao\*, <sup>2</sup>Boyu Song\*, <sup>2</sup>Shangbang Long\*, <sup>1</sup>Minghang He, <sup>3</sup>Cong Yao, <sup>1</sup>Xiang Bai  
<sup>1</sup>Huazhong University of Science and Technology, <sup>2</sup>Peking University, <sup>3</sup>Megvii Inc.  
{mhliao, minghanghe09, xbai}@hust.edu.cn  
{boyusong, shangbang.long}@pku.edu.cn, yaocong2010@gmail.com

## 国外同行举报CVPR 2020中国论文一稿多投, 作者之一华科教授还是本届CVPR领域主席

△Shangbang Long个人主页

他给出了两种猜测: 一是Shangbang Long和Cong Yao一开始和其他作者一同工作, 并投稿《中国科学》, 离开北大到CMU后又两人又对之前的工作做了改进, 重新向CVPR投稿。

另一种可能是作者希望通过改动一些实验方法和细节, 重复投稿给知名顶会, 并侥幸认为在国际并不知名《中国科学》上的同一篇论文不会其他同行被发现, 这样的问题就十分严重了。

这位网友已经向作者发邮件询问, 但没有得到明确答复, 作者只是说目前正计划向学界详细说明这些差异。但坚持说这两篇论文代表了两种不同的使用UE生成数据的尝试, 是完全不同的。

## 论文作者回应

论文作者Shangbang Long在Reddit热帖下面做出了回应, 解释了这两篇论文为什么看起来高度相似, 以及它们的不同之处。

# 案例分析

## Case6: 共同作者也存在风险

### A Roadmap for Big Model \*

Sha Yuan<sup>\*1</sup> Hanyu Zhao<sup>\*1</sup> Shuai Zhao<sup>\*1</sup> Jiahong Leng<sup>\*1</sup> Yangxiao Liang<sup>\*1</sup> Xiaozhi Wang<sup>\*2</sup> Jifan Yu<sup>\*2</sup> Xin Lv<sup>\*2</sup> Zhou Shao<sup>\*1</sup> Jiaao He<sup>\*2</sup> Yankai Lin<sup>\*3</sup> Xu Han<sup>\*2</sup> Zhenghao Liu<sup>\*4</sup> Ning Ding<sup>\*2</sup> Yongming Rao<sup>\*2</sup> Yizhao Gao<sup>\*5</sup> Liang Zhang<sup>\*5</sup> Ming Ding<sup>\*2</sup> Cong Fang<sup>\*6</sup> Yisen Wang<sup>\*6</sup> Mingsheng Long<sup>\*2</sup> Jing Zhang<sup>\*5</sup> Yinpeng Dong<sup>\*2</sup> Tianyu Pang<sup>\*2</sup> Peng Cui<sup>\*2</sup> Lingxiao Huang<sup>\*7</sup> Zheng Liang<sup>\*2</sup> Huawei Shen<sup>\*8</sup> Hui Zhang<sup>\*2</sup> Quanshi Zhang<sup>\*9</sup> Qingxiu Dong<sup>\*6</sup> Zhixing Tan<sup>\*2</sup> Mingxuan Wang<sup>\*13</sup> Shuo Wang<sup>\*2</sup> Long Zhou<sup>\*14</sup> Haoran Li<sup>\*10</sup> Junwei Bao<sup>\*10</sup> Yingwei Pan<sup>\*10</sup> Weinan Zhang<sup>\*11</sup> Zhou Yu<sup>\*12</sup> Rui Yan<sup>\*5</sup> Chence Shi<sup>\*15</sup> Minghao Xu<sup>\*15</sup> Zuobai Zhang<sup>\*15</sup> Guoqiang Wang<sup>\*1</sup> Xiang Pan<sup>\*16</sup> Mengjie Li<sup>\*17</sup> Xiaoyu Chu<sup>\*1</sup> Zijun Yao<sup>\*2</sup> Fangwei Zhu<sup>\*2</sup> Shulin Cao<sup>\*2</sup> Weicheng Xue<sup>\*2</sup> Zixuan Ma<sup>\*2</sup> Zhengyan Zhang<sup>\*2</sup> Shengding Hu<sup>\*2</sup> Yujia Qin<sup>\*2</sup> Chaojun Xiao<sup>\*2</sup> Zheni Zeng<sup>\*2</sup> Ganqu Cui<sup>\*2</sup> Weize Chen<sup>\*2</sup> Weilin Zhao<sup>\*2</sup> Yuan Yao<sup>\*2</sup> Peng Li<sup>\*3</sup> Wenzhao Zheng<sup>\*2</sup> Wenliang Zhao<sup>\*2</sup> Ziyi Wang<sup>\*2</sup> Borui Zhang<sup>\*2</sup> Nanvi Fei<sup>\*5</sup> Anwen Hu<sup>\*5</sup> Zenan Ling<sup>\*6</sup> Haoyang Li<sup>\*5</sup> Boxi Cao<sup>\*18</sup> Xianpei Han<sup>\*18</sup> Weidong Zhan<sup>\*6</sup> Baobao Chang<sup>\*6</sup> Hao Sun<sup>\*2</sup> Jiawen Deng<sup>\*2</sup> Chujie Zheng<sup>\*2</sup> Juanzi Li<sup>\*22</sup> Lei Hou<sup>\*22</sup> Xigang Cao<sup>\*21</sup> Jidong Zhai<sup>\*22</sup> Zhiyuan Liu<sup>\*22</sup> Maosong Sun<sup>\*22</sup> Jiwen Lu<sup>\*22</sup> Zhiwu Lu<sup>\*22</sup> Qin Jin<sup>\*22</sup> Ruihua Song<sup>\*5</sup> Ji-Rong Wen<sup>\*5</sup> Zhouchen Lin<sup>\*22</sup> Liwei Wang<sup>\*22</sup> Hang Su<sup>\*22</sup> Jun Zhu<sup>\*22</sup> Zhifang Sui<sup>\*22</sup> Jiajun Zhang<sup>\*22</sup> Yang Liu<sup>\*22</sup> Xiaodong He<sup>\*22</sup> Minlie Huang<sup>\*22</sup> Jian Tang<sup>\*22</sup> **Jie Tang<sup>\*22,1</sup>**

- <sup>1</sup> Beijing Academy of Artificial Intelligence
- <sup>2</sup> Tsinghua University
- <sup>3</sup> Wechat, Tencent Inc.
- <sup>4</sup> Northeastern University
- <sup>5</sup> Renmin University of China
- <sup>6</sup> Peking University
- <sup>7</sup> Huawei TCS Lab
- <sup>8</sup> Institute of Computing Technology, Chinese Academy of Sciences
- <sup>9</sup> Shanghai Jiao Tong University
- <sup>10</sup> JD AI Research
- <sup>11</sup> Harbin Institute of Technology
- <sup>12</sup> Columbia University
- <sup>13</sup> ByteDance AI Lab
- <sup>14</sup> Microsoft Research Asia
- <sup>15</sup> Mila-Quebec AI Institute & University of Montreal
- <sup>16</sup> New York University
- <sup>17</sup> BeiHang University
- <sup>18</sup> Institute of Software, Chinese Academy of Sciences
- <sup>19</sup> Institute of Automation Chinese Academy of Sciences

Jie Tang designs this big model roadmap. Authors labeled with \* organize different parts of this report. Authors labeled with \* contribute equally. They are ranked according to their section.

知乎 首页 知乎知学堂 等你来答 知乎直答 专家谈重庆人工增雨引八级

提问 消息 私信 创作中心

抄襲 人工智能 清华大学 北京智源人工智能研究院 论文抄襲

关注者 3,559 被浏览 8,194,861

### 如何看待智源、清华等单位论文 A Roadmap for Big Model 中大量段落被指涉嫌抄襲?

Google Brain研究员Nicholas Carlini近日在一篇博客中指出智源、清华等单位的论文A Roadmap for Big Mod...显示全部

已关注 写回答 邀请回答 好问题 591 37 条评论 分享

匿名用户

776 人赞同了该回答

不幸出现在作者列表里的学术民工，来写点参与这篇文章的一些体验。智源已经要求不要自行回应了，所以请大家别问也别猜个人身份。

- 整个项目是唐杰发起的，牵头干活的是袁莎，title应该是大模型研究中心副主任还是什么，总之是唐杰在智源的心腹。其实我一直没搞明白袁莎这个人有什么成就，个人感觉不像认真搞过学术的人。这点从她给整篇文章写的那个潦草的写作规范就能看出来，比如说要求引用统一用Google Scholar导出的bib，了解的人应该知道这有多离谱，说实在的您哪怕用dblp呢。
- 智源高层之间的内斗是很有意思的。所以大家还真以为这事肯定处理不到唐杰，智源领导里跟唐杰有矛盾的可不止一个。
- 老板把这个活派给我的时候，留给我的时间只有一周多。我不太清楚这是我老板的问题还是智源组织方面的问题，反正写不出来那就是我的问题。可能有的老板派给学生的时间更紧，有人选择直接抄也就可以理解了。当然可以理解不代表可以原谅，我的每一个字都是自己写的，看到这个消息心里只有愤怒。
- 在2021年12月上旬交稿之后不到一周，我们整个组访问这篇文章Overleaf Project的权限就被撤掉了。从那直到arxiv前，我再也没有机会看到这篇文章哪怕一眼。作为共同作者我应该在公开前起码看一遍，但智源并没给这个机会。2022年3月底，这篇文章突然就挂arxiv了。理论上中间这四个月袁莎这些智源的人是应该校对、修改、统一风格的，但看上去她们就是拼一起就发了，typo都没改。

Xiv:2203.14101v3 [cs.LG] 2 Apr 2022

## Case7: 聚焦于你的自己的工作

```
@GetMapping("/exam/{examCode}")
public ApiResult findById(@PathVariable("examCode") Integer examCode){
    System.out.println("根据 ID 查找");
    ExamManage res = examManageService.findById(examCode);
    if(res == null) {
        return ApiResultHandler.buildApiResult(10000,"考试编号不存在",null);
    }
    return ApiResultHandler.buildApiResult(200,"请求成功!",res);
}
```

(该部分是对试卷进行增删改查操作的功能代码, 重复文献: 王怀\_基于 Java Web 技术的易库在线考试系统的设计与实现, 可能会存在代码结构和变量名重复的可能)

表 4-8 学生信息表的结构

列名	数据类型	长度	主键否	可空否	功能描述
<u>Studentid</u>	Varchar	20	是	否	学生学号
<u>studentName</u>	varchar	20	否	是	学生名字
Grade	varchar	10	否	是	年级

Fig.: Past source code might increase duplicate rate.

## Case8: 聚焦于你的自己的工作

### 2.2 Spring Boot 技术

Spring 框架是 Java 平台上一种应用框架，Spring 框架具有面向切面编程 (AOP) 框架，Spring AOP 框架是基于代理模式的，同时运行时可配置，AOP 框架则主要是针对模块之间的交叉关注点进行模块化。基于 Spring 的优秀特性，设计出了一种 Spring Boot 框架。Spring boot 框架中有两个我们非常实用的策略，第一个是开箱即用 (Outofbox)，在开发过程中，通过在 MAVEN 项目的 pom 文件中添加相关依赖包，然后使用对应注解来代替繁琐的 XML 配置文件。（重复文献是 201600114221\_翟志斌\_大学本科毕业论文 翟志斌-《大学生》重复原

**不要过于详细地介绍流行的工具**

# 案例分析

## Case9: 因论文抄袭而不能正常毕业

- 不是人文学科
  - 经常要讨论经典论述而导致文字重复率高
- 计算机学科：正常重复率应该远小于 5%
  - 查重率低不是没有抄袭的充分条件
- 查重率明显小于学校要求不意味着合格
- 查重只是预警，确认学术不端面临严重后果

文字拷贝后自己都极难发现！  
更不能指望导师。

毕业论文(1)



作者:

提交时间 2022-04-20 12:40:14

总相似比 6.31%

自写率 93.69%

复写率 6.31%

自引率 0.0%

引用率 0.0%

总字数 10255

重复字数 438

读万卷书 行万里路 只为最好的修炼  
论文是为研究工作服务的  
研究是为产业落地服务的  
所有的研究动机都来源于个人的内生动力



QQ: 14777591 (宇宙骑士)  
Email: [ouxinyu@alumni.hust.edu.cn](mailto:ouxinyu@alumni.hust.edu.cn)  
Website: <http://ouxinyu.cn>  
Tel: 18687840023