大心作业 1-3 以及期末论文

October 8-9, 2024

Zhicheng Lin (linzc@ustc.edu.cn)

总体目的：培养文献检索、分析能力，了解最新研究，锻炼批判性思维和创新能力

总体任务：从 1994 年到 2023 年，每个学生负责某一年份 Nature 或者 Science 杂志（不含子刊）：

1）寻找**所有与“人”相关的文章，包括心理学、社会科学、神经科学（如大脑研究）等领域**，但不包括生物医学（比如人类胎儿、器官等，因为这些不是直接针对人的层面）
2）下载这些文章
3）对这些文章进行分析、拓展（后续期末论文，之后会说明）

Science 指导

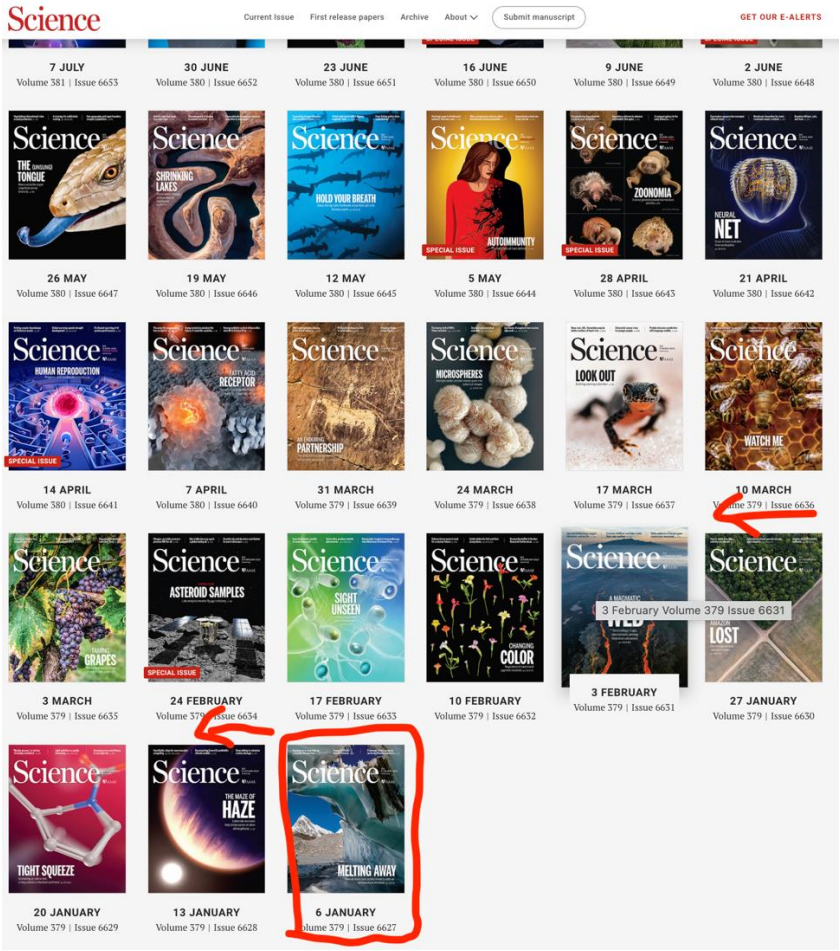
1）期刊以及年份

Science 1880 创刊以来的文章都在这里：
https://www.science.org/loi/science
可以看出，从 1994 年到 2023 年，每年有 4 期（Volume），每期包含 12/13/14 本（Issue），每年共 51 本（Issue）。

按照上面链接找对应的年份（请先**通过学校网址登陆**，如第 4 页所示——这样就可以查看和下载文章全文，否则只能查看标题和摘要）。

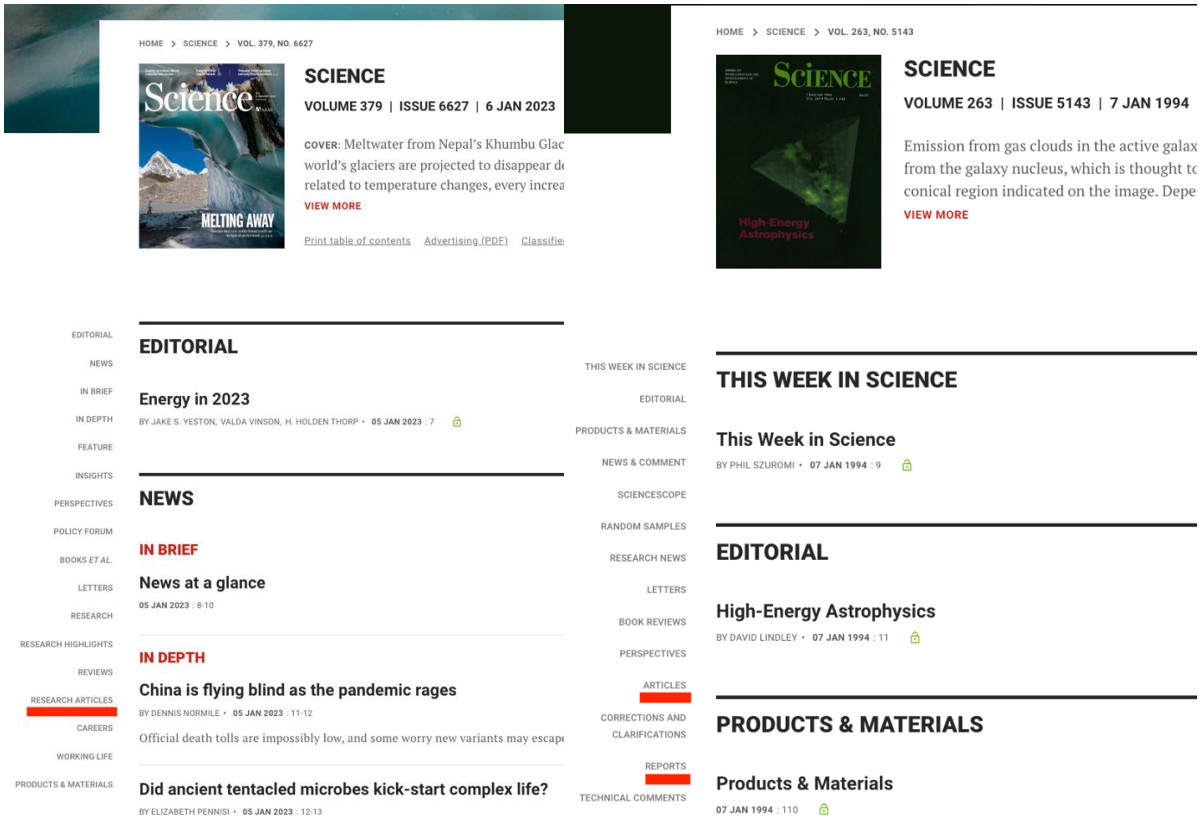
2）每一年文章寻找顺序：按从年前到年后、日期从早到晚的顺序

按照**最右下边**开始，**逆序**，如下图：



3）文章判断：

每本（Issue）有很多不同类似的文章，比如 Editorial、News 等等，我们只看研究类型的文章，这个不同年份会有些改变，但是我们研究文章主要是：**Research Articles、Articles、Reports**——可以只看这三个类型（如有不确定之处，请联系我们）。在这些类型的文章里，每篇文章如何判断是否属于人的研究请继续阅读。



根据文章的**标题 (Title，一般涉及文章的主题和发现)、摘要 (Abstract，一般涉及文章的研究背景、对象、问题、方法、结果和意义)** 判断是不是有可能涉及到人，**包括心理学、社会科学、神经科学（brain）等研究**，但**不包括生物医学研究（比如人类胎儿、人类器官等，因为不是直接在人的层面上）**。如英文不确定也可以用翻译软件（如 ChatGPT 等大语言模型、DeepL 等）。

不确定的话看**全文**（具体如何看全文，请看下面说明）里面的**方法 (Method)**。

还不确定的话，把文章链接发到**群里咨询**。

例子：

比如下面左边这篇虽然与人有关，但主要是讨论人类活动对雪的影响，不是关于人的行为、心理、社会、神经活动等的研究，所以不算。类似地，右侧的文章虽然也提到了人，但主要是关于一种新算法/软件，即 AlphaGeometry 怎么证明奥赛几何题的研究。

Article | Open access | Published: 10 January 2024

Evidence of human influence on Northern Hemisphere snow loss

Alexander R. Gottlieb & Justin S. Mankin

Nature **625**, 293–300 (2024) | Cite this article

39k Accesses | **14** Citations | **1607** Altmetric | Metrics

Abstract

Documenting the rate, magnitude and causes of snow loss is essential to benchmark the pace of climate change and to manage the differential water security risks of snowpack declines[1,2,3,4]. So far, however, observational uncertainties in snow mass[5,6] have made the detection and attribution of human-forced snow losses elusive, undermining societal preparedness. Here we show that human-caused warming has caused declines in Northern Hemisphere-scale March snowpack over the 1981–2020 period. Using an ensemble of snowpack reconstructions, we identify robust snow trends in 82 out of 169 major Northern Hemisphere river basins, 31 of which we can confidently attribute to human influence. Most crucially, we show a generalizable and highly nonlinear temperature sensitivity of snowpack, in which snow becomes marginally more sensitive to one degree Celsius of warming as climatological winter temperatures exceed minus eight degrees Celsius. Such nonlinearity explains the lack of widespread snow loss so far and augurs much sharper declines and water security risks in the most populous basins. Together, our results emphasize that human-forced snow losses and their water consequences are attributable—even absent their clear detection in individual snow products—and will accelerate and homogenize with near-term warming, posing risks to water resources in the absence of substantial climate mitigation.

Article | Open access | Published: 17 January 2024

Solving olympiad geometry without human demonstrations

Trieu H. Trinh, Yuhuai Wu, Quoc V. Le, He He & Thang Luong

Nature **625**, 476–482 (2024) | Cite this article

265k Accesses | **28** Citations | **1022** Altmetric | Metrics

An Author Correction to this article was published on 23 February 2024

This article has been updated

Abstract

Proving mathematical theorems at the olympiad level represents a notable milestone in human-level automated reasoning[1,2,3,4], owing to their reputed difficulty among the world's best talents in pre-university mathematics. Current machine-learning approaches, however, are not applicable to most mathematical domains owing to the high cost of translating human proofs into machine-verifiable format. The problem is even worse for geometry because of its unique translation challenges[1,5], resulting in severe scarcity of training data. We propose AlphaGeometry, a theorem prover for Euclidean plane geometry that sidesteps the need for human demonstrations by synthesizing millions of theorems and proofs across different levels of complexity. AlphaGeometry is a neuro-symbolic system that uses a neural language model, trained from scratch on our large-scale synthetic data, to guide a symbolic deduction engine through infinite branching points in challenging problems. On a test set of 30 latest olympiad-level problems, AlphaGeometry solves 25, outperforming the previous best method that only solves ten problems and approaching the performance of an average International Mathematical Olympiad (IMO) gold medallist. Notably, AlphaGeometry produces human-readable proofs, solves all geometry problems in the IMO 2000 and 2015 under human expert evaluation and discovers a generalized version of a translated IMO theorem in 2004.

https://www.nature.com/articles/s41586-023-06794-y

下面左边是关于 Twitter 用户的研究，所以算。下面右边虽然主要是关于语言模型的，但是因为和人做比较，所以也算。

Analysis | Open access | Published: 02 October 2024

Differences in misinformation sharing can lead to politically asymmetric sanctions

Mohsen Mosleh, Qi Yang, Tauhid Zaman, Gordon Pennycook & David G. Rand

Nature (2024) | Cite this article

10k Accesses | **393** Altmetric | Metrics

Abstract

In response to intense pressure, technology companies have enacted policies to combat misinformation[1,2,3,4]. The enforcement of these policies has, however, led to technology companies being regularly accused of political bias[5,6,7]. We argue that differential sharing of misinformation by people identifying with different political groups[8,9,10,11,12,13,14,15] could lead to political asymmetries in enforcement, even by unbiased policies. We first analysed 9,000 politically active Twitter users during the US 2020 presidential election. Although users estimated to be pro-Trump/conservative were indeed substantially more likely to be suspended than those estimated to be pro-Biden/liberal, users who were pro-Trump/conservative also shared far more links to various sets of low-quality news sites—even when news quality was determined by politically balanced groups of laypeople, or groups of only Republican laypeople—and had higher estimated likelihoods of being bots. We find similar associations between stated or inferred conservatism and low-quality news sharing (on the basis of both expert and politically balanced layperson ratings) in 7 other datasets of sharing from Twitter, Facebook and survey experiments, spanning 2016 to 2023 and including data from 16 different countries. Thus, even under politically neutral anti-misinformation policies, political asymmetries in enforcement should be expected. Political imbalance in enforcement need not imply bias on the part of social media companies implementing anti-misinformation policies.

Article | Open access | Published: 25 September 2024

Larger and more instructable language models become less reliable

Lexin Zhou, Wout Schellaert, Fernando Martínez-Plumed, Yael Moros-Daval, Cèsar Ferri & José Hernández-Orallo

Nature **634**, 61–68 (2024) | Cite this article

42k Accesses | **1** Citations | **428** Altmetric | Metrics

Abstract

The prevailing methods to make large language models more powerful and amenable have been based on continuous scaling up (that is, increasing their size, data volume and computational resources[1]) and bespoke shaping up (including post-filtering[2,3], fine tuning or use of human feedback[4,5]). However, larger and more instructable large language models may have become less reliable. By studying the relationship between difficulty concordance, task avoidance and prompting stability of several language model families, here we show that easy instances for human participants are also easy for the models, but scaled-up, shaped-up models do not secure areas of low difficulty in which either the model does not err or human supervision can spot the errors. We also find that early models often avoid user questions but scaled-up, shaped-up models tend to give an apparently sensible yet wrong answer much more often, including errors on difficult questions that human supervisors frequently overlook. Moreover, we observe that stability to different natural phrasings of the same question is improved by scaling-up and shaping-up interventions, but pockets of variability persist across difficulty levels. These findings highlight the need for a fundamental shift in the design and development of general-purpose artificial intelligence, particularly in high-stakes areas for which a predictable distribution of errors is paramount.

https://www.nature.com/articles/s41586-024-07942-8
https://www.nature.com/articles/s41586-024-07930-y

4）登记文章到 excel，并下载全文

对于判断确实属于人的研究，请：

- 从**原文复制你判断的依据**，比如：here we show that easy instances for human participants

are also easy for the models, but scaled-up, shaped-up models do not secure areas of low difficulty in which either the model does not err or human supervision can spot the errors.

- 在 **excel 模版填写文章相关信息**，包括：杂志、年份、期 (Volume)、文章题目、文章链接、判断依据、是否存疑 (是 1; 否 0)
- **下载全文**，具体操作如下图所示：

https://i.ustc.edu.cn/ -》校外应用 -》Science-》登陆学校账号