Guest talk: Research dissemination

Dr. Nataliia Stulova

KSE, Kyiv, Ukraine
25.06.2025

LibreOffice

$ whoami

BSc in Systems Analysis
MSc in Artificial Intelligence
PhD in Software, Sytems and Computing

Post-docs in EPFL and UniBe
Staff Research Scientist at MacPaw

Computer scientist working on tool and techniques to understand what the programs are actually doing

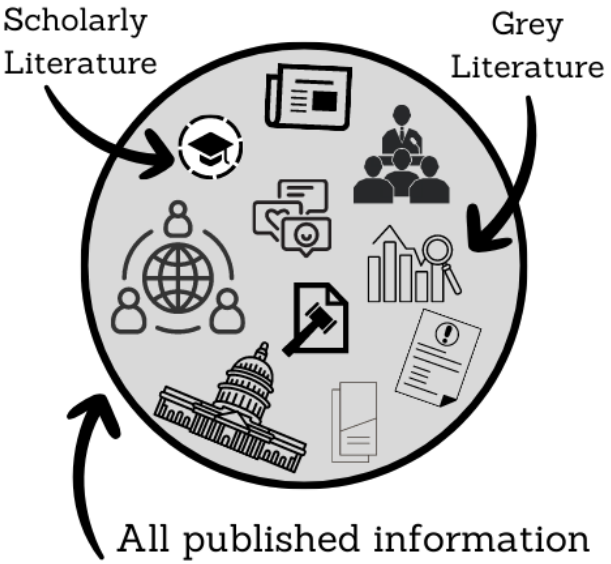
Trust nobody, question everything

so, research dissemination:

- non-research tech publications
- conferences, journals, books
- incremental publication flow
- money: APC and predatory publishing
- science is [not beyond] politics
- bonus: fun + Q&A

Illustrations by Pixeltrue on icons8

Image, text: University of Illinois at Urbana-Champaign
https://guides.library.illinois.edu/c.php?g=1310347



Grey VS white* literature

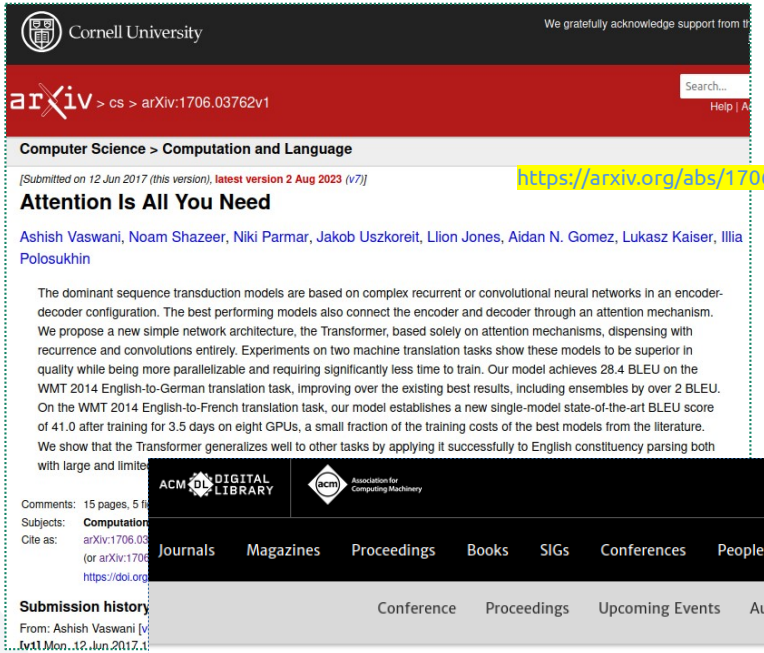
***funny how we keep "white = good" in research**

> *Grey Literature is a category that includes media, resources, documents, data, etc. that was not produced by traditional academic or commercial publishing systems, which is often referred to as White Literature.*
>
> ***Unlike White Literature, Grey Literature is not peer reviewed and is not typically published in books or scholarly journals.***

non-research tech publications

Why grey literature?

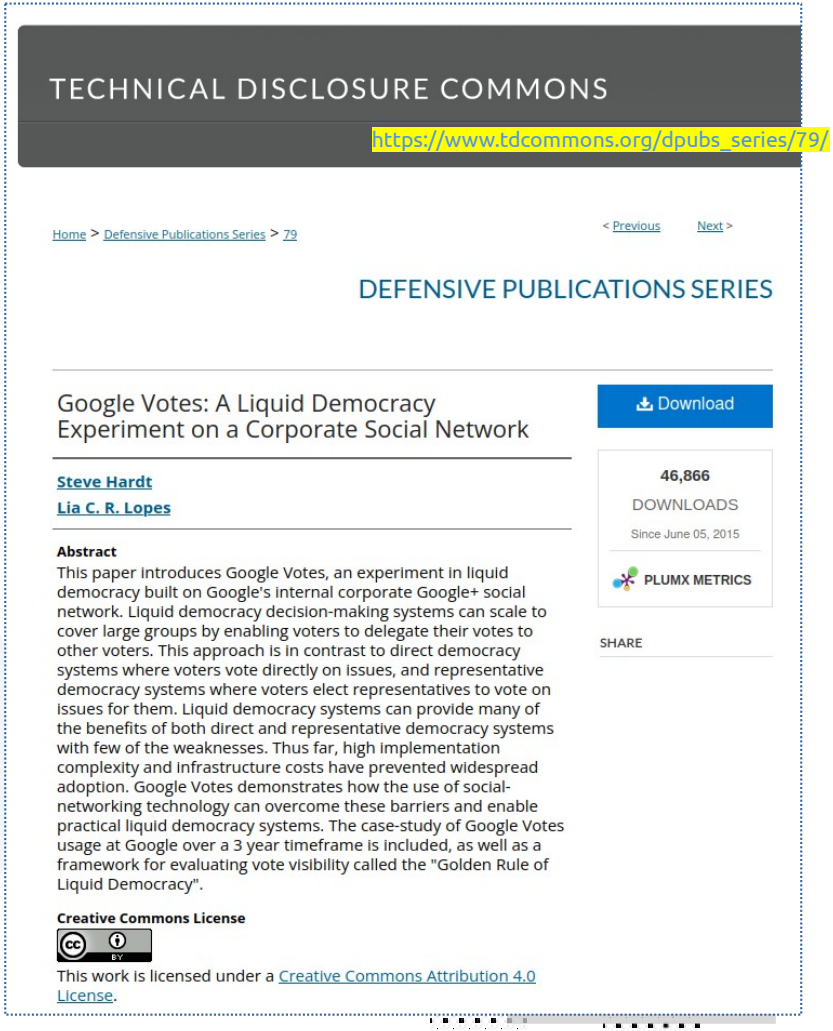
- When speed is more important for IP
- Archival practices for intermediate stage work
- Company policies (PR, branding)



non-research tech publications

Why grey literature?

- When speed is more important for IP
- Archival practices for intermediate stage work
- Company policies (PR, branding)



TECHNICAL DISCLOSURE COMMONS

https://www.tdcommons.org/dpubs_series/79/

Home > Defensive Publications Series > 79

< Previous Next >

DEFENSIVE PUBLICATIONS SERIES

Google Votes: A Liquid Democracy Experiment on a Corporate Social Network

Steve Hardt

Lia C. R. Lopes

Download

46,866

DOWNLOADS

Since June 05, 2015

PLUMX METRICS

SHARE

Abstract

This paper introduces Google Votes, an experiment in liquid democracy built on Google's internal corporate Google+ social network. Liquid democracy decision-making systems can scale to cover large groups by enabling voters to delegate their votes to other voters. This approach is in contrast to direct democracy systems where voters vote directly on issues, and representative democracy systems where voters elect representatives to vote on issues for them. Liquid democracy systems can provide many of the benefits of both direct and representative democracy systems with few of the weaknesses. Thus far, high implementation complexity and infrastructure costs have prevented widespread adoption. Google Votes demonstrates how the use of social-networking technology can overcome these barriers and enable practical liquid democracy systems. The case-study of Google Votes usage at Google over a 3 year timeframe is included, as well as a framework for evaluating vote visibility called the "Golden Rule of Liquid Democracy".

Creative Commons License

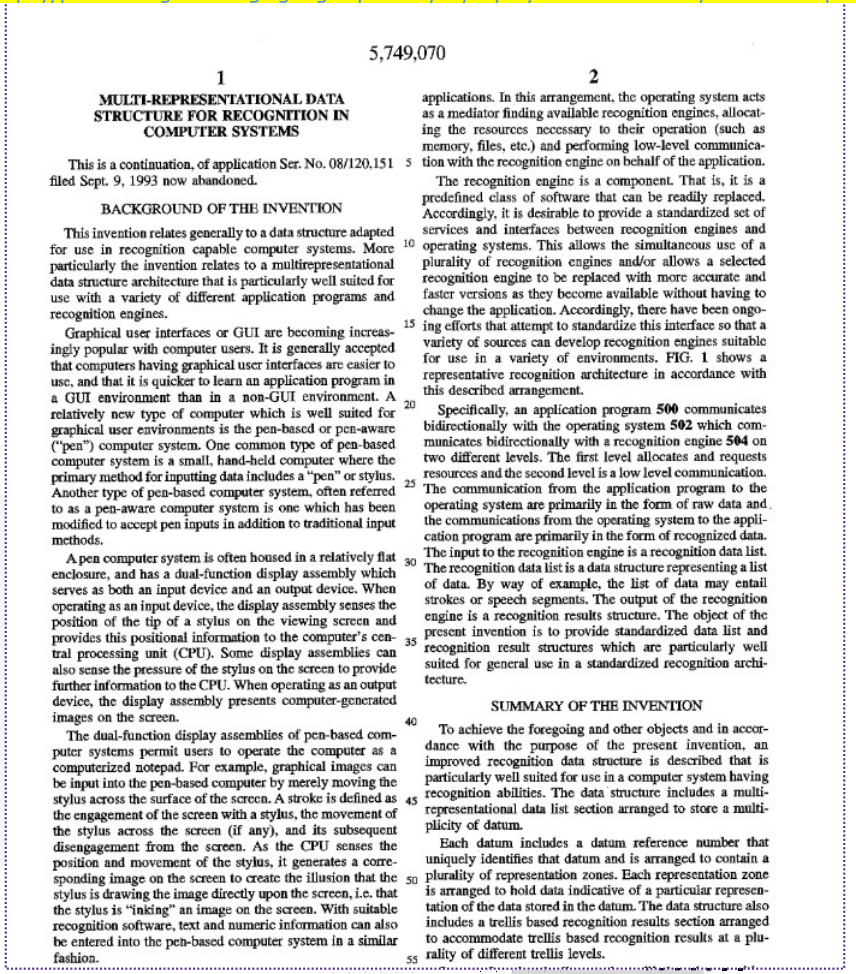
This work is licensed under a Creative Commons Attribution 4.0 License.

non-research tech publications

Why grey literature?

- When speed is more important for IP
- Archival practices for intermediate stage work
- Company policies (PR, branding)

https://patentimages.storage.googleapis.com/9d/82/b8/644ca1555f3521/US5749070.pdf



5,749,070

1

MULTI-REPRESENTATIONAL DATA STRUCTURE FOR RECOGNITION IN COMPUTER SYSTEMS

This is a continuation, of application Ser. No. 08/120,151 filed Sept. 9, 1993 now abandoned.

BACKGROUND OF THE INVENTION

This invention relates generally to a data structure adapted for use in recognition capable computer systems. More particularly the invention relates to a multirepresentational data structure architecture that is particularly well suited for use with a variety of different application programs and recognition engines.

Graphical user interfaces or GUI are becoming increasingly popular with computer users. It is generally accepted that computers having graphical user interfaces are easier to use, and that it is quicker to learn an application program in a GUI environment than in a non-GUI environment. A relatively new type of computer which is well suited for graphical user environments is the pen-based or pen-aware ("pen") computer system. One common type of pen-based computer system is a small, hand-held computer where the primary method for inputting data includes a "pen" or stylus. Another type of pen-based computer system, often referred to as a pen-aware computer system is one which has been modified to accept pen inputs in addition to traditional input methods.

A pen computer system is often housed in a relatively flat enclosure, and has a dual-function display assembly which serves as both an input device and an output device. When operating as an input device, the display assembly senses the position of the tip of a stylus on the viewing screen and provides this positional information to the computer's central processing unit (CPU). Some display assemblies can also sense the pressure of the stylus on the screen to provide further information to the CPU. When operating as an output device, the display assembly presents computer-generated images on the screen.

The dual-function display assemblies of pen-based computer systems permit users to operate the computer as a computerized notepad. For example, graphical images can be input into the pen-based computer by merely moving the stylus across the surface of the screen. A stroke is defined as the engagement of the screen with a stylus, the movement of the stylus across the screen (if any), and its subsequent disengagement from the screen. As the CPU senses the position and movement of the stylus, it generates a corresponding image on the screen to create the illusion that the stylus is drawing the image directly upon the screen, i.e. that the stylus is "inking" an image on the screen. With suitable recognition software, text and numeric information can also be entered into the pen-based computer system in a similar fashion.

2

applications. In this arrangement, the operating system acts as a mediator finding available recognition engines, allocating the resources necessary to their operation (such as memory, files, etc.) and performing low-level communication with the recognition engine on behalf of the application.

The recognition engine is a component. That is, it is a predefined class of software that can be readily replaced. Accordingly, it is desirable to provide a standardized set of services and interfaces between recognition engines and operating systems. This allows the simultaneous use of a plurality of recognition engines and/or allows a selected recognition engine to be replaced with more accurate and faster versions as they become available without having to change the application. Accordingly, there have been ongoing efforts that attempt to standardize this interface so that a variety of sources can develop recognition engines suitable for use in a variety of environments. FIG. 1 shows a representative recognition architecture in accordance with this described arrangement.

Specifically, an application program **500** communicates bidirectionally with the operating system **502** which communicates bidirectionally with a recognition engine **504** on two different levels. The first level allocates and requests resources and the second level is a low level communication. The communication from the application program to the operating system are primarily in the form of raw data and the communications from the operating system to the application program are primarily in the form of recognized data. The input to the recognition engine is a recognition data list. The recognition data list is a data structure representing a list of data. By way of example, the list of data may entail strokes or speech segments. The output of the recognition engine is a recognition results structure. The object of the present invention is to provide standardized data list and recognition result structures which are particularly well suited for general use in a standardized recognition architecture.

SUMMARY OF THE INVENTION

To achieve the foregoing and other objects and in accordance with the purpose of the present invention, an improved recognition data structure is described that is particularly well suited for use in a computer system having recognition abilities. The data structure includes a multi-representational data list section arranged to store a multiplicity of datum.

Each datum includes a datum reference number that uniquely identifies that datum and is arranged to contain a plurality of representation zones. Each representation zone is arranged to hold data indicative of a particular representation of the data stored in the datum. The data structure also includes a trellis based recognition results section arranged to accommodate trellis based recognition results at a plurality of different trellis levels.

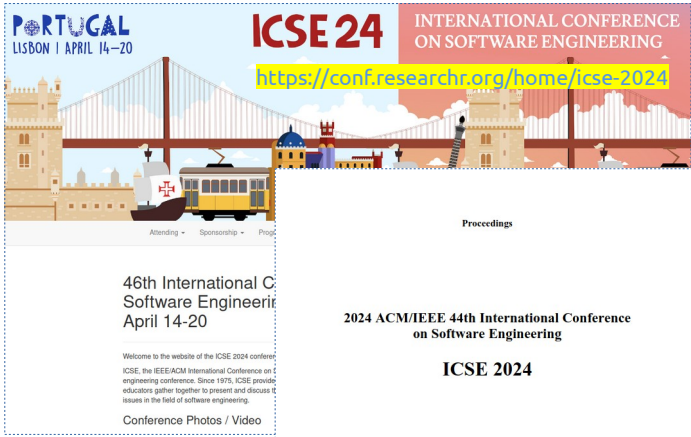
non-research tech publications


Image credits: Warner Bros Pictures

From grey to white: peer review

Quality control from domain experts

> **Conferences**: 2-4 reviews per submission from the Program Committee members, 1-2 review rounds (optional rebuttal), time deadline, contributions from everyone



https://dl.acm.org/doi/proceedings/10.1145/3597503

conferences, journals, books

Image credits: Warner Bros Pictures

From grey to white: peer review

Quality control from domain experts

Conferences: 2-4 reviews per submission from the Program Committee members, 1-2 review rounds (optional rebuttal), time deadline, contributions from everyone

Journals: 2-4 reviews per manuscript from individually invited domain experts, several revision iterations until reviewer consensus, issue format, contributions from everyone

https://dl.acm.org/toc/tosem/2025/34/5

conferences, journals, books

Image credits: Warner Bros Pictures

From grey to white: peer review

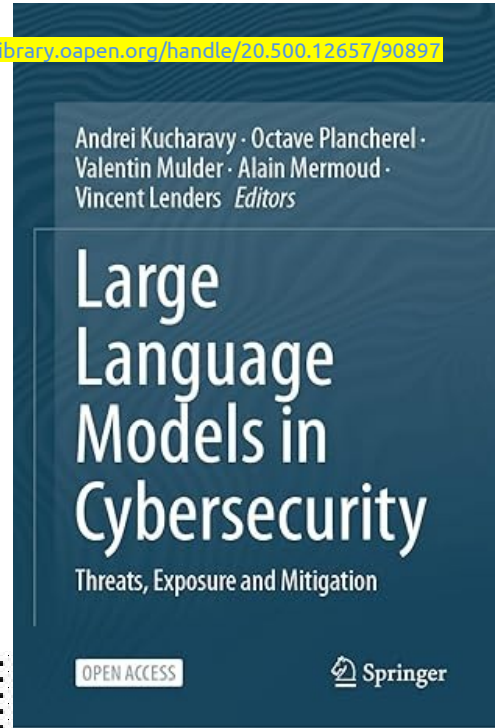
Quality control from domain experts

Conferences: 2-4 reviews per submission from the Program Committee members, 1-2 review rounds (optional rebuttal), time deadline, contributions from everyone

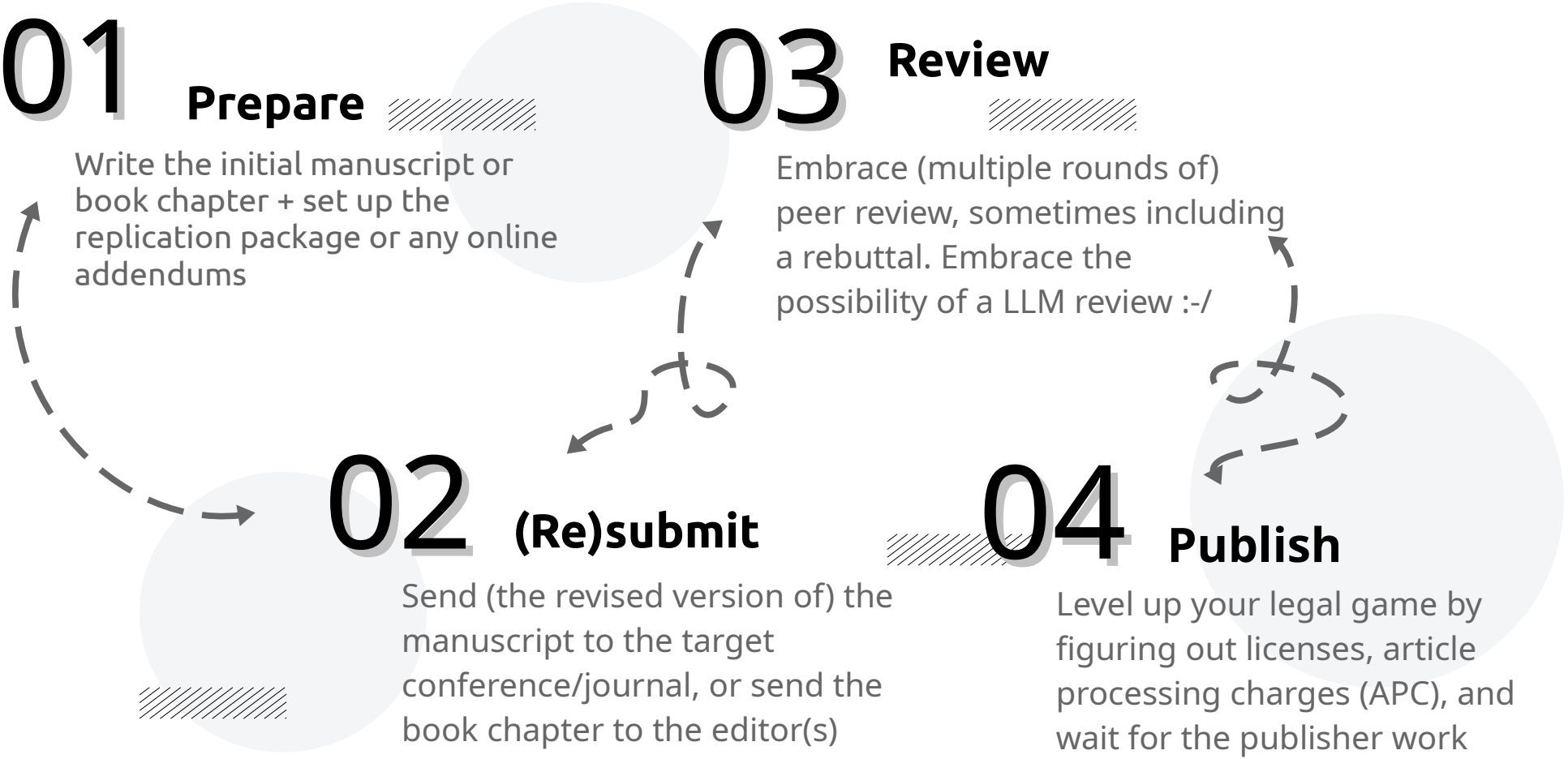
Journals: 2-4 reviews per manuscript from individually invited domain experts, several revision iterations until reviewer consensus, issue format, contributions from everyone

Books: Editors' responsibility, invitation-based contribution in chapter format from well-established field experts

https://library.oapen.org/handle/20.500.12657/90897



conferences, journals, books



incremental publication flow

incremental publication flow

Extended Abstract / Poster

The smallest possible research paper. EA can range from 1-2 to 8 pages, the poster itself is typicaly A0 format and depends on your taste and design skills.

Position or vision paper

Less formal than a regular short paper, presents an opinion, can be technical too

ERA, NIER, RENE, Data, Tools, Industry paper

Specialized short papers to present ideas in the initial development state, demo prototypes, do replication studies,or describe real-world use

Short paper

Format one would use for reporting preliminary research results and gain community feedback, 4-5 pages limit

Full/journal-first paper

Most comonly used format in Computer Science conferences, 8-12 pages, strict page limit (for review fairness)

Journal paper

Does not really have a fixed length limit anymore, often is created from a full conference paper + 30% or new content (experimens, proofs, data etc.)

Academic Publishing Business

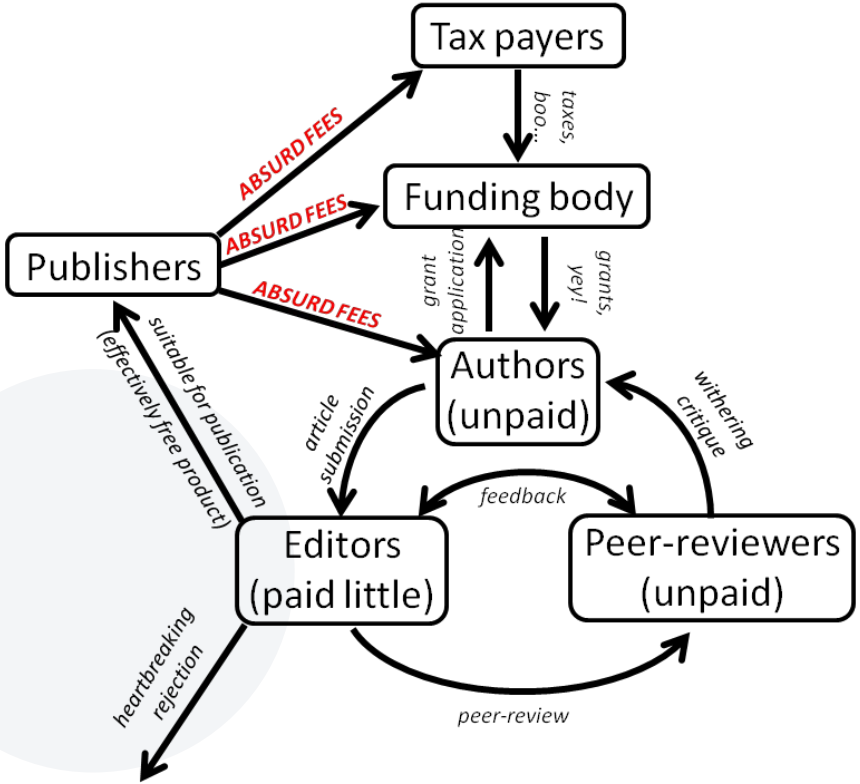


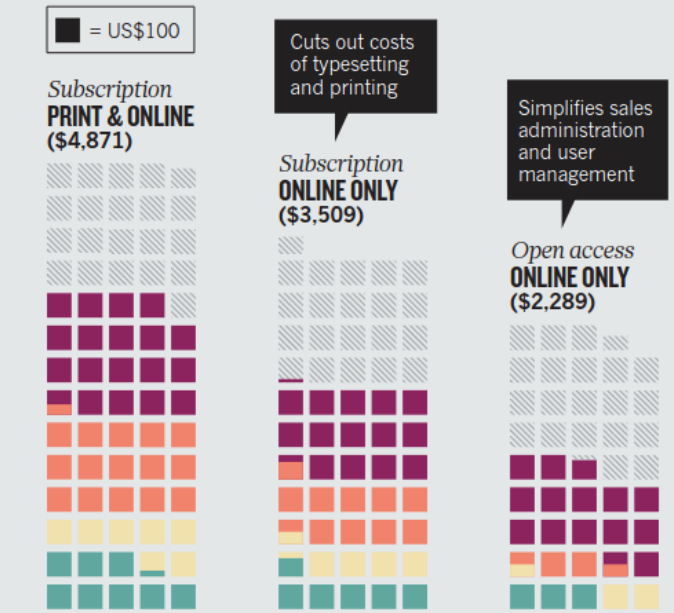
Image: https://katatrepsis.com/2012/01/28/scientific-publishing-what-a-con/

Default: paid individual access, free access via a library, authors tranfer copyright to the publisher but pay no APC

Alternatve: open-access publishing, where authors retain the copyright, pay the APC or wait out an embargo period, but the publication is free to access to everyone

How costs break down

An economic model shows how switching from subscription to open access changes the costs of publishing.



Voluntary peer review (not counted in price)
Additional cost if reviewers were paid for their time.

Article processing
Administering peer review (assuming average rejection rate of 50%); editing; proofreading; typesetting; graphics; quality assurance.

Other costs
Covers, indexes and editorial; rights management; sales and payments; printing and delivery; online user management; marketing and communications; helpdesk; online hosting.

Management and investment
Includes cost to establish journal: assumed 20% subscription; 15% open access.

Margin
Assumed 20% subscription; 15% open access.

https://www.nature.com/articles/495426a

Data from J. Houghton *et al. Economic implications of alternative scholarly publishing models* (Joint Information Systems Committee, 2009). available at go.nature.com/uqrxqw.

Default: paid individual access, free access via a library, authors tranfer copyright to the publisher but pay no APC

Alternatve: open-access publishing, where authors retain the copyright, pay the APC or wait out an embargo period, but the publication is free to access to everyone

“Publish or perish”

Requirements on number of publications motivate people to artificially inflate publication numbres

Publisher “About” information is absent on the journal’s web site

Publisher direct marketing (i.e., spamming) or other advertising is obtrusive

Instructions to authors information is not available

Information on peer review and copyright is absent or unclear on the journal web site

Journal scope statement is absent or extremely vague

No information is provided about the publisher, or the information provided does not clearly indicate a relationship to a mission to disseminate research content

Repeat lead authors in same issue

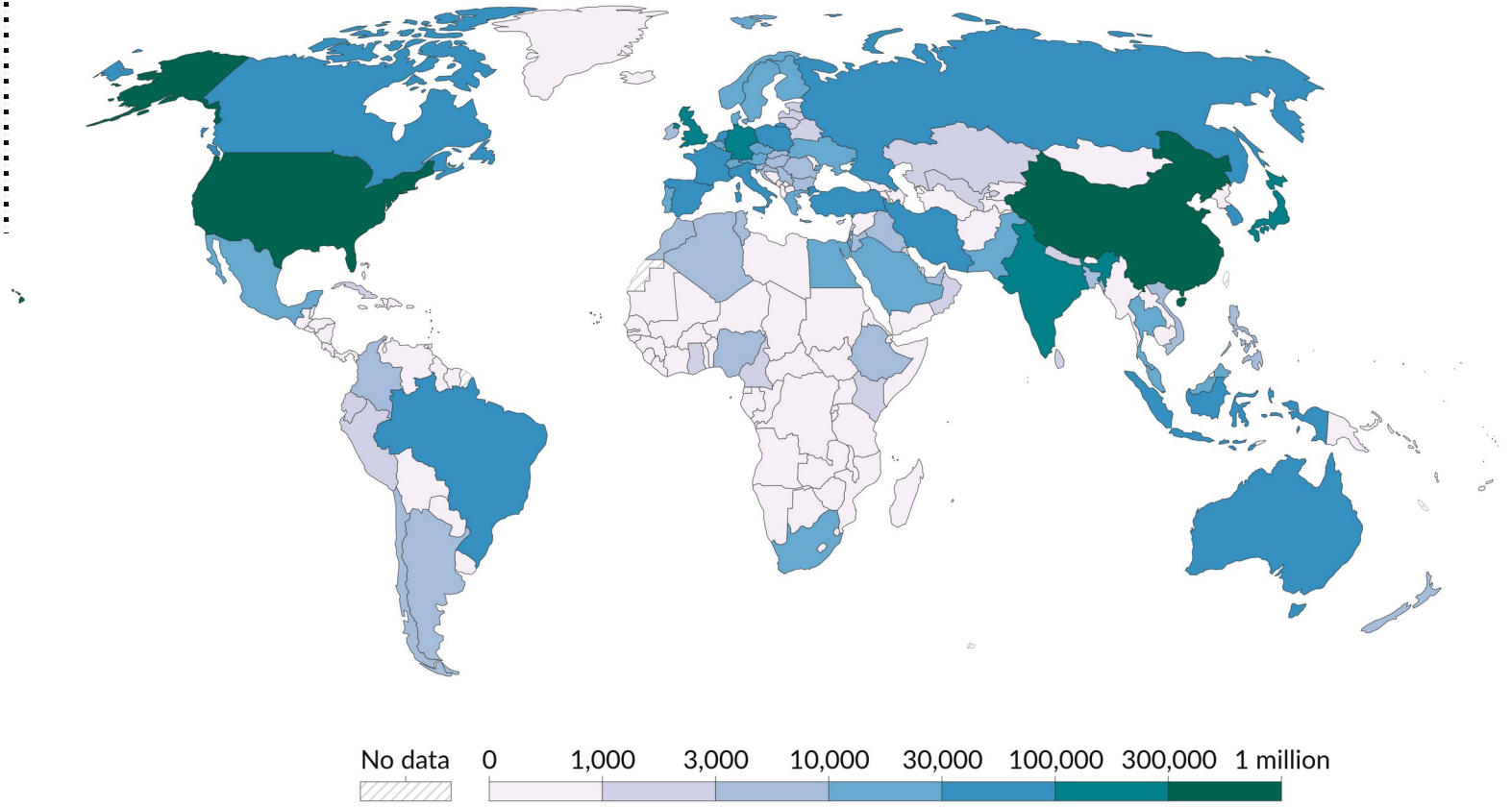
Publisher has a negative reputation (e.g., documented examples in Chronicle of Higher Education, list-servs, etc.)

Image: https://umlibguides.um.edu.my/c.php?g=938956&p=6796235

Annual articles published in scientific and technical journals, 2020

Includes physics, biology, chemistry, mathematics, clinical medicine, biomedical research, engineering and technology, and earth and space sciences.

Our World in Data



Data source: National Science Foundation, via World Bank (2025)

OurWorldinData.org/research-and-development | CC BY

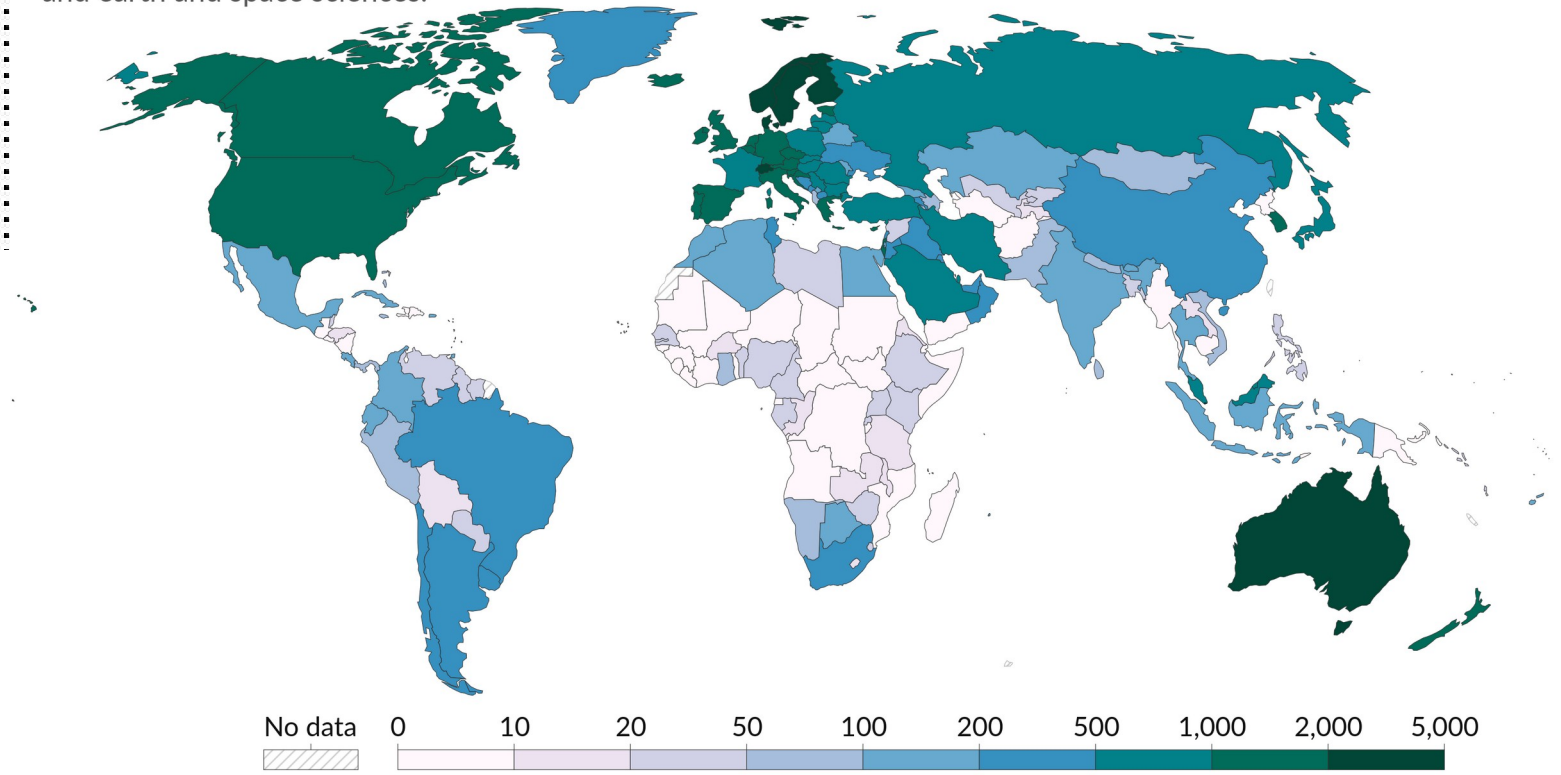
Note: Articles are counted by the country of the author's institution.

science is [not beyond] politics

Annual articles published in scientific and technical journals per million people, 2020

Our World in Data

Includes physics, biology, chemistry, mathematics, clinical medicine, biomedical research, engineering and technology, and earth and space sciences.



Data source: National Science Foundation, via World Bank (2025); United Nations Population Division, Eurostat, national statistical offices, and United Nations Statistics Division, via World Bank (2025)

Note: Articles are assigned based on the country of the first author's institution.

OurWorldinData.org/research-and-development | CC BY

science is [not beyond] politics

https://orbilu.uni.lu/bitstream/10993/60055/1/AndroZooARetrospectiveWithAGlimpseIntoTheFuture.pdf

AndroZoo: A Retrospective with a Glimpse into the Future

Marco Alecci
SnT, University of Luxembourg,
Luxembourg, Luxembourg
marco.alecci@uni.lu

Pedro Jesús Ruiz Jiménez
SnT, University of Luxembourg,
Luxembourg, Luxembourg
pedro.ruiz@uni.lu

Kevin Allix
Independent researcher,
Rennes, France
kallix@kallix.net

Tegawendé F. Bissyandé
SnT, University of Luxembourg,
Luxembourg, Luxembourg
tegawende.bissyande@uni.lu

Jacques Klein
SnT, University of Luxembourg,
Luxembourg, Luxembourg
jacques.klein@uni.lu

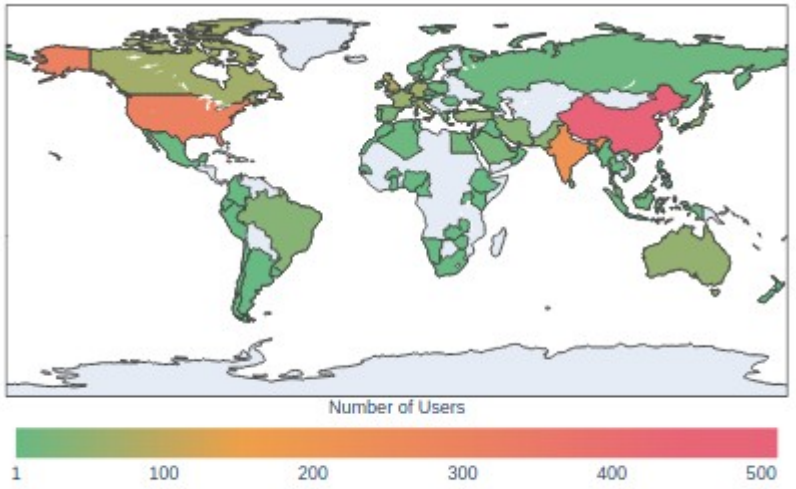
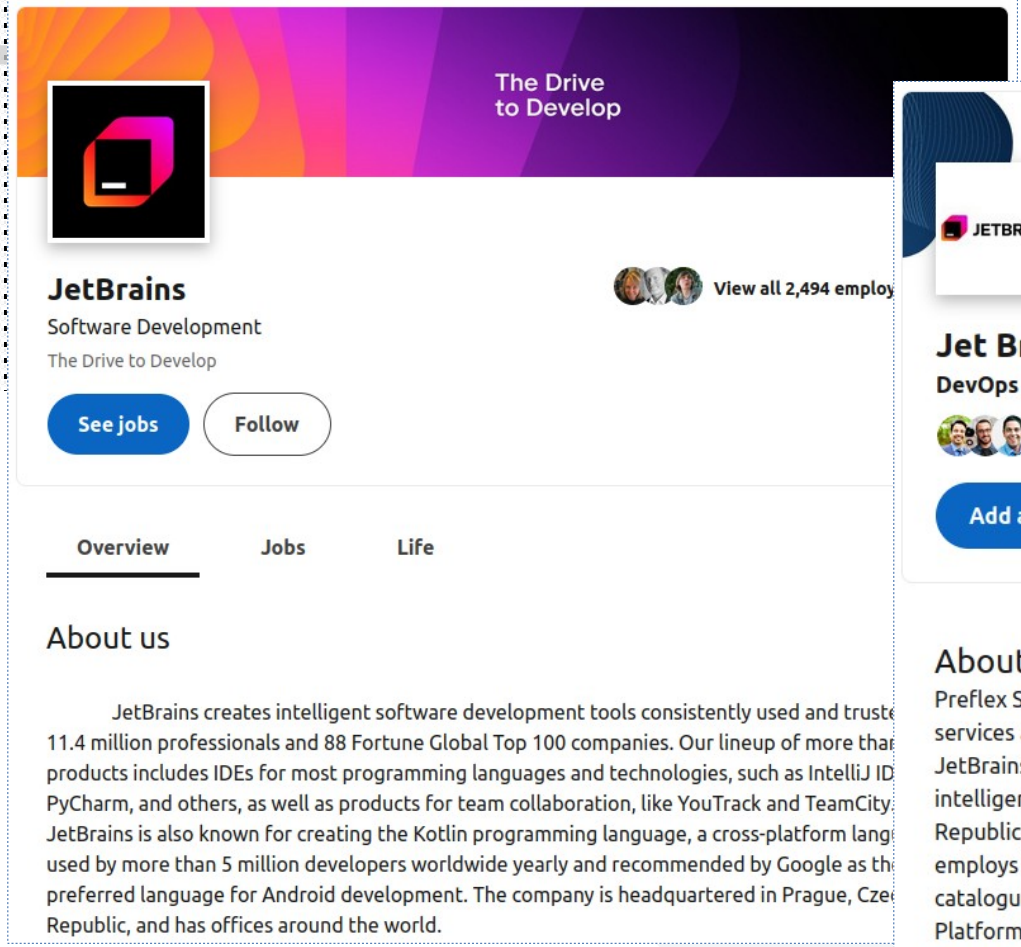
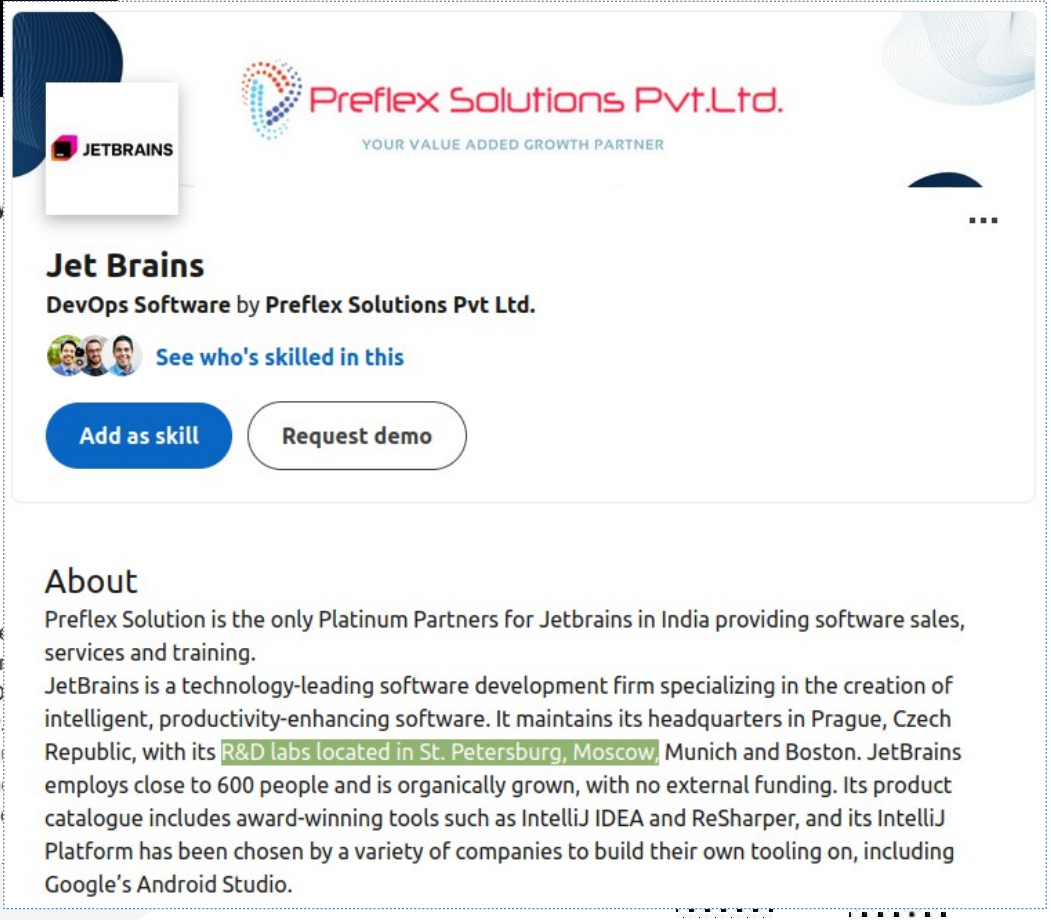


Figure 3: User Distribution across the world

science is [not beyond] politics





17:15 5m Talk ☆ **RefExpo: Unveiling Software Project Structures through Advanced Dependency Graph Extraction**
MSR Data and Tool Showcase Track
Vahid Haratian Bilkent Univeristy, Pouria Derakhshanfar JetBrains Research, Vladimir Kovalenko JetBrains Research, Eray Tüzün Bilkent University

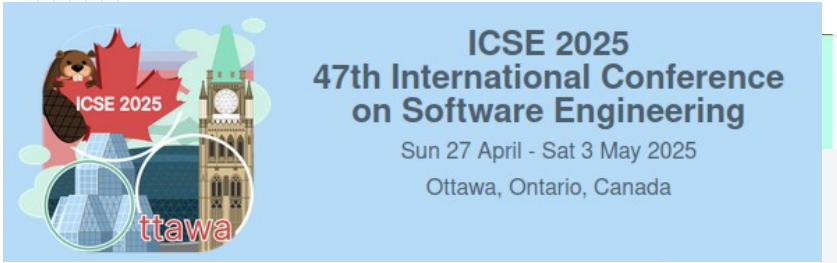
16:30 5m Talk ☆ **Drawing Pandas: A Benchmark for LLMs in Generating Plotting Code**
MSR Data and Tool Showcase Track
Timur Galimzyanov JetBrains Research, Sergey Titov JetBrains Research, Yaroslav Golubev JetBrains Research, Egor Bogomolov JetBrains Research
Pre-print

14:00 - 15:30 **FORGE2025 Keynote at 207**
Chair(s): Michele Tufano Google

14:00 45m Keynote ☆ **Industry Keynote: One shall not live on LLM alone**
Darya Rovdo JetBrains

16:00 - 17:30 **Process at 203** New Ideas and Emerging Results (NIER) / Journal-first Papers / Research Track / SE In Practice (SEIP)
Chair(s): Luigi Benedicenti University of New Brunswick

16:00 15m Talk ☆ **Full Line Code Completion: Bringing AI to Desktop**
SE In Practice (SEIP)
Anton Semenkin JetBrains, Vitaliy Bibaev JetBrains, Yaroslav Sokolov JetBrains, Kirill Krylov JetBrains, Alexey Kalina JetBrains, Anna Khannanova JetBrains, Danila Savenkov JetBrains, Darya Rovdo JetBrains, Igor Davidenko JetBrains, Kirill Karnaukhov JetBrains, Maxim Vakhrushev JetBrains, Mikhail Kostyukov JetBrains, Mikhail Podvitskii JetBrains, Petr Surkov JetBrains, Yaroslav Golubev JetBrains Research, Nikita Povarov JetBrains, Timofey Bryksin JetBrains Research
Pre-print



16:30 15m Talk ☆ **Towards Realistic Evaluation of Commit Message Generation by Matching Online and Offline Settings**
SE In Practice (SEIP)
Petr Tsvetkov JetBrains Research, Aleksandra Eliseeva JetBrains Research, Danny Dig University of Colorado Boulder, JetBrains Research, Alexander Bezzubov JetBrains, Yaroslav Golubev JetBrains Research, Timofey Bryksin JetBrains Research, Yaroslav Zharov JetBrains Research
Pre-print

09:00 - 10:30 **Opening & Keynotes at 205**
Chair(s): Danny Dig University of Colorado Boulder, JetBrains Research

09:00 10m Day opening ☆ **Welcoming Message by Organizers**
Danny Dig University of Colorado Boulder, JetBrains Research

09:10 40m Keynote ☆ **Time Travel with IDEs: From Line Editors to AI Copilots**
Dave Thomas Bedarra Corp

09:50 40m Keynote ☆ **AI Agents Inside the IDE: Possibilities and Practice**
Yaroslav Zharov JetBrains Research

11:00 - 12:30 **Session 1: Visions of the future, & Session 2: Diversity of environments at 205** IDE
Chair(s): Carolin Brandt TU Delft, Danny Dig University of Colorado Boulder, JetBrains Research
11:00–11:45 — Session 1. Visions of the future.
11:45–12:30 — Session 2. Diversity of environments.

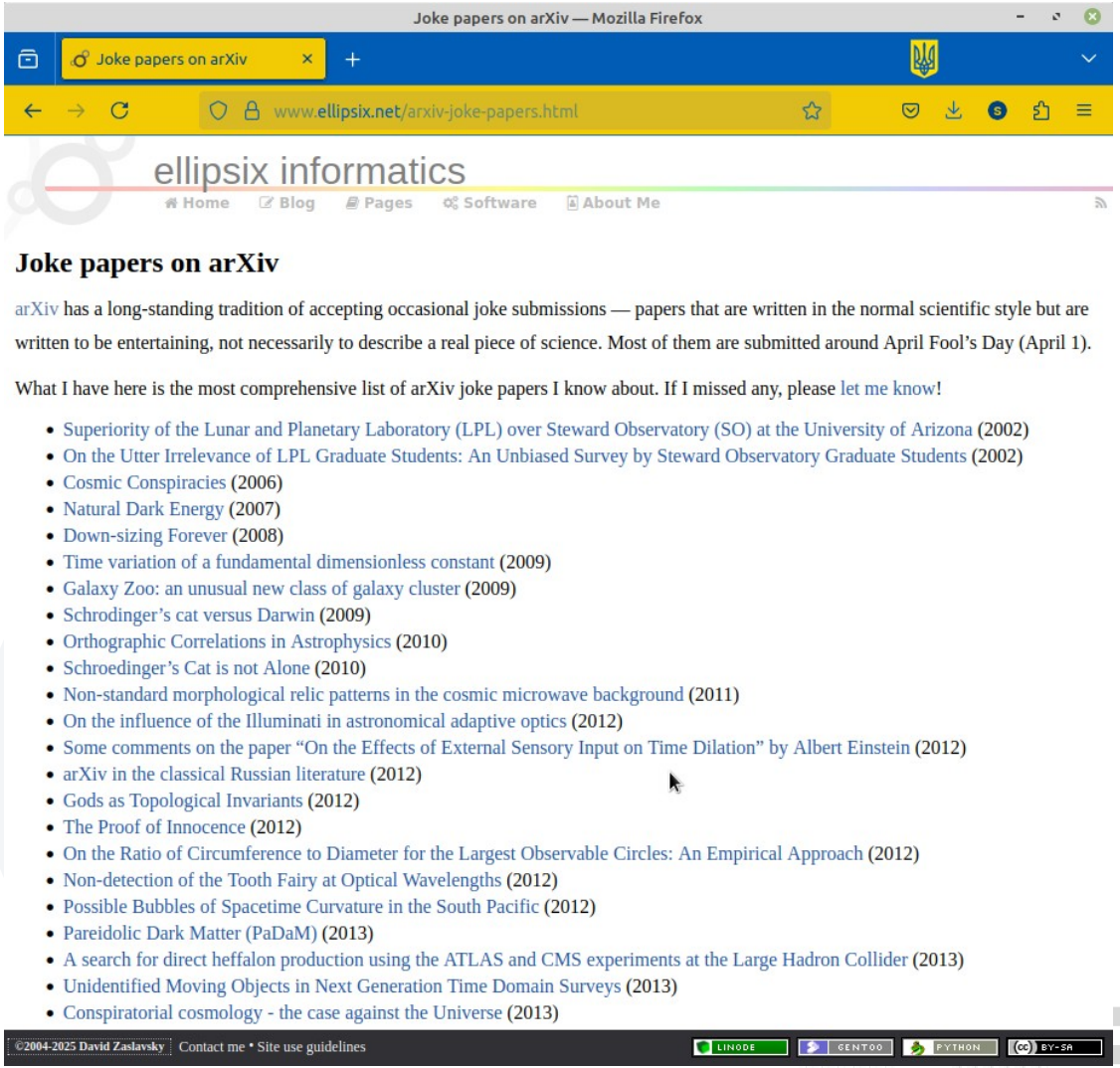
11:00 15m Talk ☆ **In-IDE Programming Courses: Learning Software Development in a Real-World Setting**
Anastasiia Birillo JetBrains Research, Ilya Vlasov JetBrains Research, Katsiaryna Dzialets JetBrains, Hieke Keuning Utrecht University, Timofey Bryksin JetBrains Research
Pre-print

12:15 15m Talk ☆ **Evolving the Computational Notebook: A Two-Dimensional Canvas for Enhanced Human-AI Interaction**
Konstantin Grotov JetBrains Research, Constructor University, Dmitry Botov Neapolis University Paphos
Pre-print

14:00 - 15:30 **Session 3: Refactoring & AI, & Session 4: Plugins and applications at 205**
Chair(s): Danny Dig University of Colorado Boulder, JetBrains Research, Darya Rovdo JetBrains
14:00–14:45 — Session 3. Refactoring and AI.
14:45–15:30 — Session 4. Plugins and applications.

16:40 10m Talk ☆ **YABLoCo: Yet Another Benchmark for Long Context Code Generation (Virtual Talk)**
Aidar Valeev Innopolis University, Vladimir Ivanov Innopolis University, Roman Garaev Innopolis University, Vadim Lomshakov JetBrains, Irina Piontkovskaya Huawei Noah's Ark Lab, Israel Adewuyi Innopolis University

science is [not beyond] politics


Joke papers on arXiv — Mozilla Firefox
Joke papers on arXiv
www.ellipsix.net/arxiv-joke-papers.html
ellipsix informatics
Home
Blog
Pages
Software
About Me
Joke papers on arXiv
arXiv has a long-standing tradition of accepting occasional joke submissions — papers that are written in the normal scientific style but are written to be entertaining, not necessarily to describe a real piece of science. Most of them are submitted around April Fool's Day (April 1).
What I have here is the most comprehensive list of arXiv joke papers I know about. If I missed any, please let me know!
Superiority of the Lunar and Planetary Laboratory (LPL) over Steward Observatory (SO) at the University of Arizona (2002)
On the Utter Irrelevance of LPL Graduate Students: An Unbiased Survey by Steward Observatory Graduate Students (2002)
Cosmic Conspiracies (2006)
Natural Dark Energy (2007)
Down-sizing Forever (2008)
Time variation of a fundamental dimensionless constant (2009)
Galaxy Zoo: an unusual new class of galaxy cluster (2009)
Schrodinger's cat versus Darwin (2009)
Orthographic Correlations in Astrophysics (2010)
Schroedinger's Cat is not Alone (2010)
Non-standard morphological relic patterns in the cosmic microwave background (2011)
On the influence of the Illuminati in astronomical adaptive optics (2012)
Some comments on the paper "On the Effects of External Sensory Input on Time Dilation" by Albert Einstein (2012)
arXiv in the classical Russian literature (2012)
Gods as Topological Invariants (2012)
The Proof of Innocence (2012)
On the Ratio of Circumference to Diameter for the Largest Observable Circles: An Empirical Approach (2012)
Non-detection of the Tooth Fairy at Optical Wavelengths (2012)
Possible Bubbles of Spacetime Curvature in the South Pacific (2012)
Pareidolic Dark Matter (PaDaM) (2013)
A search for direct heffalon production using the ATLAS and CMS experiments at the Large Hadron Collider (2013)
Unidentified Moving Objects in Next Generation Time Domain Surveys (2013)
Conspiratorial cosmology - the case against the Universe (2013)
©2004-2025 David Zaslavsky
Contact me • Site use guidelines
LINODE
GENTOO
PYTHON
BY-SA