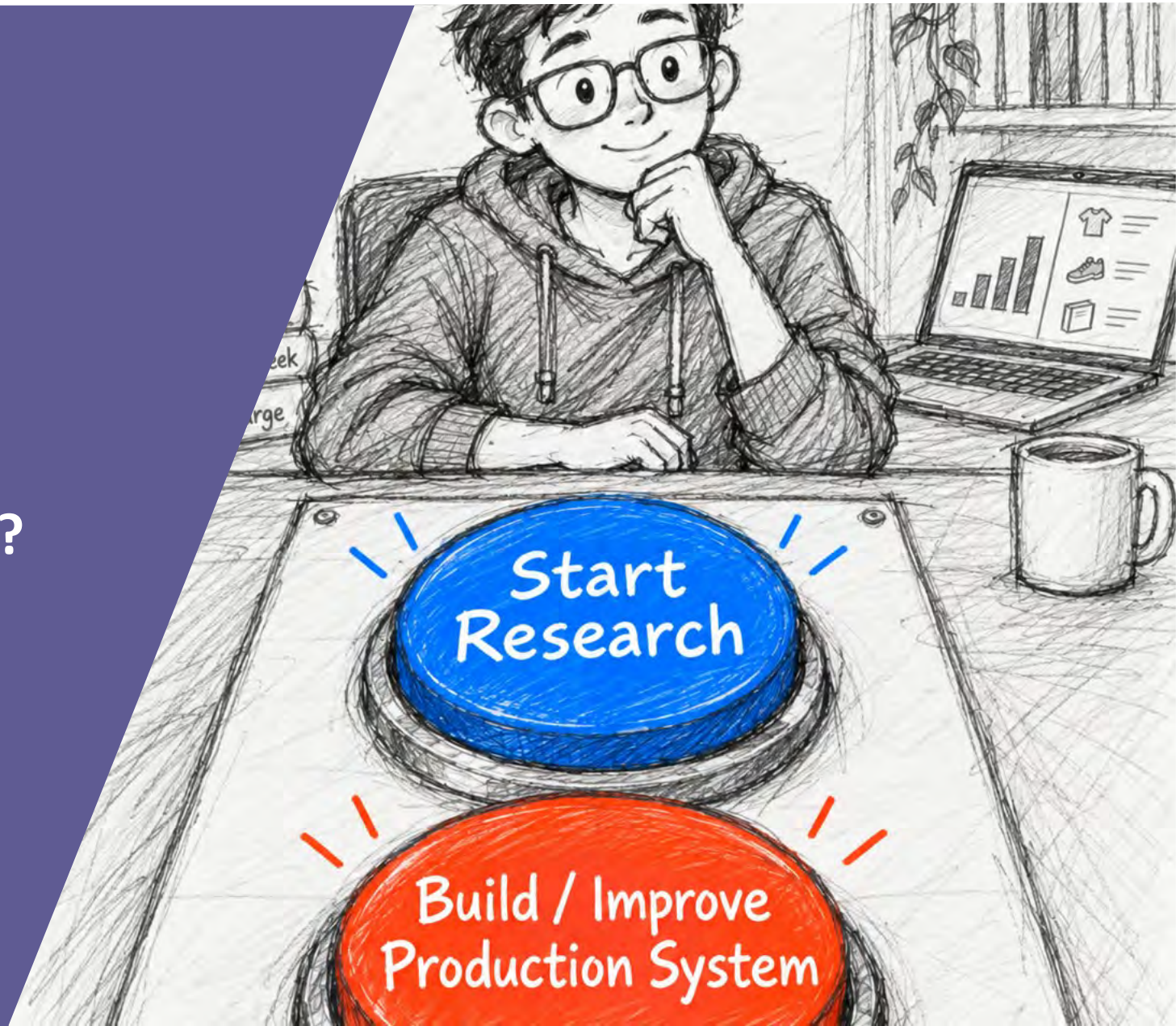
Universität Siegen / Naturwissenschaftlich-Technische Fakultät / Intelligent Systems Group

When AI Does RecSys Experiments for $1

Why Spend Four Years on a PhD?

Joeran Beel, University of Siegen

27th Sept. 2026, Doctoral Consortium at ACM RecSys 2026



Content

1. Today‘s AI Scientists
2. Tomorrow‘s Autonomous RecSys Research Labs
3. The Future of today‘s PhD students

Joeran Beel

University of Siegen

Professor of Intelligent Systems

DAAD German Academic Exchange Service

Scientific Advisory Board Member for Artificial Intelligence at the German Academic Exchange Service (DAAD)

Recommender-Systems.com

Founder

acm

Outreach Editor and AE of ACM TORS
General Co-Chair of ACM RecSys 2027 (Hawaii)
Member of the ACM RecSys Steering Committee

Former Affiliations

University of California, Berkeley

Trinity College Dublin

NII National Institute of Informatics, Tokyo

beel.org

gscholar.beel.org

@jbeel

@JoeranBeel

1. Today‘s AI Scientists

https://pdos.csail.mit.edu/archive/scigen/

SciGen (2005)

Rooter: A Methodology for the Typical Unification of Access Points and Redundancy

Jeremy Stribling, Daniel Aguayo and Maxwell Krohn

ABSTRACT

Many physicists would agree that, had it not been for congestion control, the evaluation of web browsers might never have occurred. In fact, few hackers worldwide would disagree with the essential unification of voice-over-IP and public-private key pair. In order to solve this riddle, we confirm that SMPs can be made stochastic, cacheable, and interposable.

I. INTRODUCTION

Many scholars would agree that, had it not been for active networks, the simulation of Lamport clocks might never have occurred. The notion that end-users synchronize with the investigation of Markov models is rarely outdated. A theoretical grand challenge in theory is the important unification

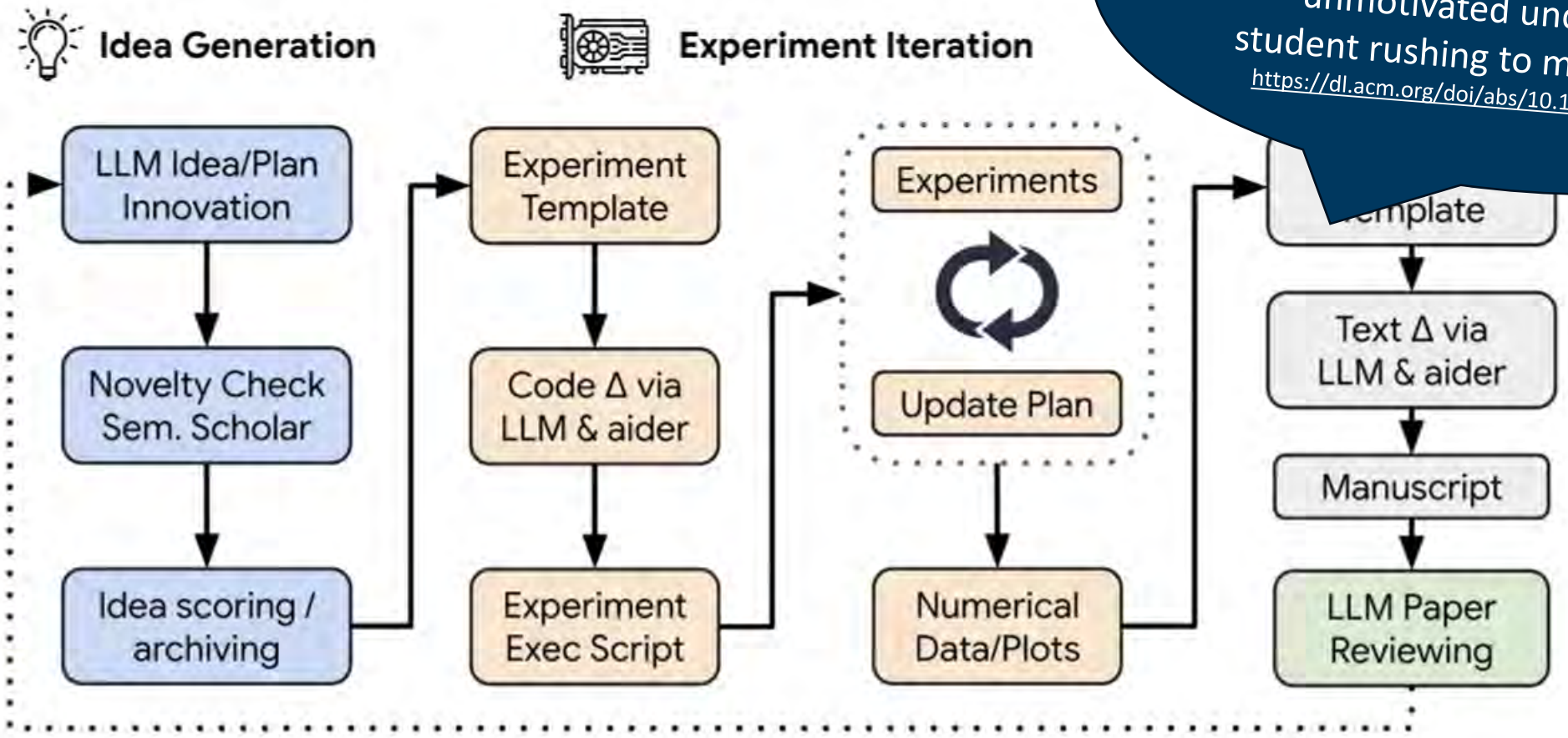
The rest of this paper is organized as follows. For starters, we motivate the need for fiber-optic cables. We place our work in context with the prior work in this area. To address this obstacle, we disprove that even though the much-tauted autonomous algorithm for the construction of digital-to-analog converters by Jones [10] is NP-complete, object-oriented languages can be made signed, decentralized, and signed. Along these same lines, to accomplish this mission, we concentrate our efforts on showing that the famous ubiquitous algorithm for the exploration of robots by Sato et al. runs in $\Omega((n + \log n))$ time [22]. In the end, we conclude.

II. ARCHITECTURE

Our research is principled. Consider the early methodology by Martin and Smith; our model is similar, but will actually

Sakana.AI‘s „AI Scientist“ (2024)

“The the quality of its manuscripts currently aligns with that of an unmotivated undergraduate student rushing to meet a deadline”
https://dl.acm.org/doi/abs/10.1145/3769733.3769747



https://arxiv.org/pdf/2408.06292

- **1st Generation: Chat (understand and communicate)**
- **2nd Generation: Code (tools)**
- **3rd Generation: Research, 2026:**
  - Google — ScientistTwo, Scientist-One, Gemini for Science
  - Meta — Auto-RecSys / Ranking Engineer Agent
  - Anthropic — Claude Science
  - Microsoft — Microsoft Discovery

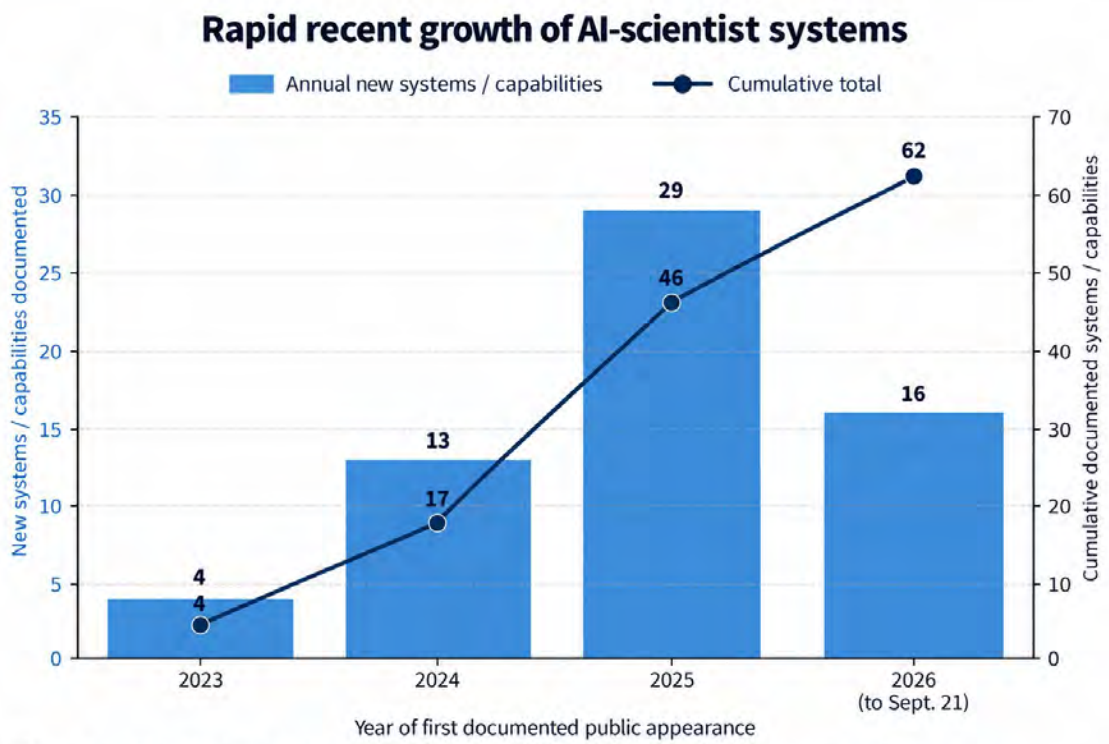


Figure. Annual introductions (bars) and cumulative total (line) of documented AI-scientist or research-agent systems/capabilities in the survey's expanded inventory. Counts are selected inventory milestones rather than all field-wide tools. 2026 is partial, covering systems documented through 21 Sept. 2026. Source: Beel, Two Decades After the Robot Scientist (v2, 22 Sept. 2026).

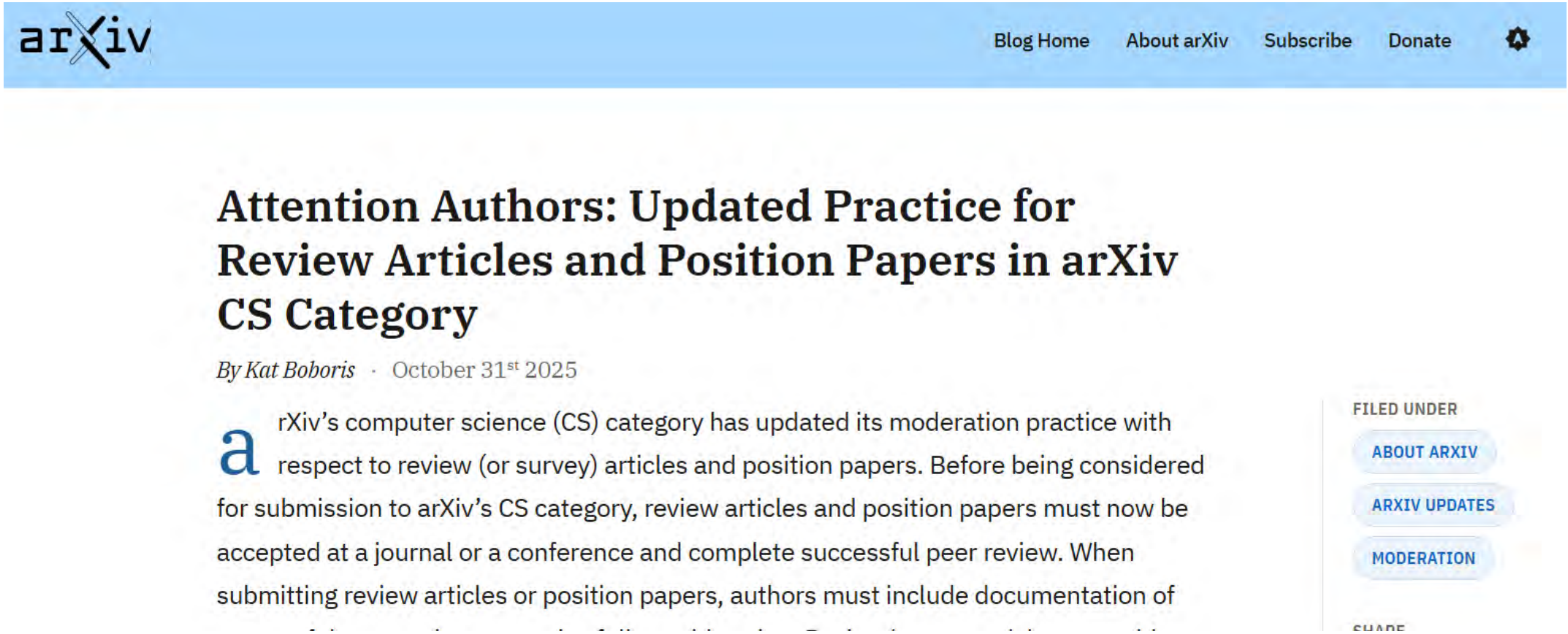
AI Generated. Based on AI Survey
https://www.researchgate.net/publication/414552473_Two_Decades_After_the_Robot_Scientist_A_Survey_of_AI_Scientists_Autonomous_Research_and_Why_Validation_Still_Sets_the_Pace

AI-Research works, it‘s fast, it‘s scalable, and it‘s cheap!

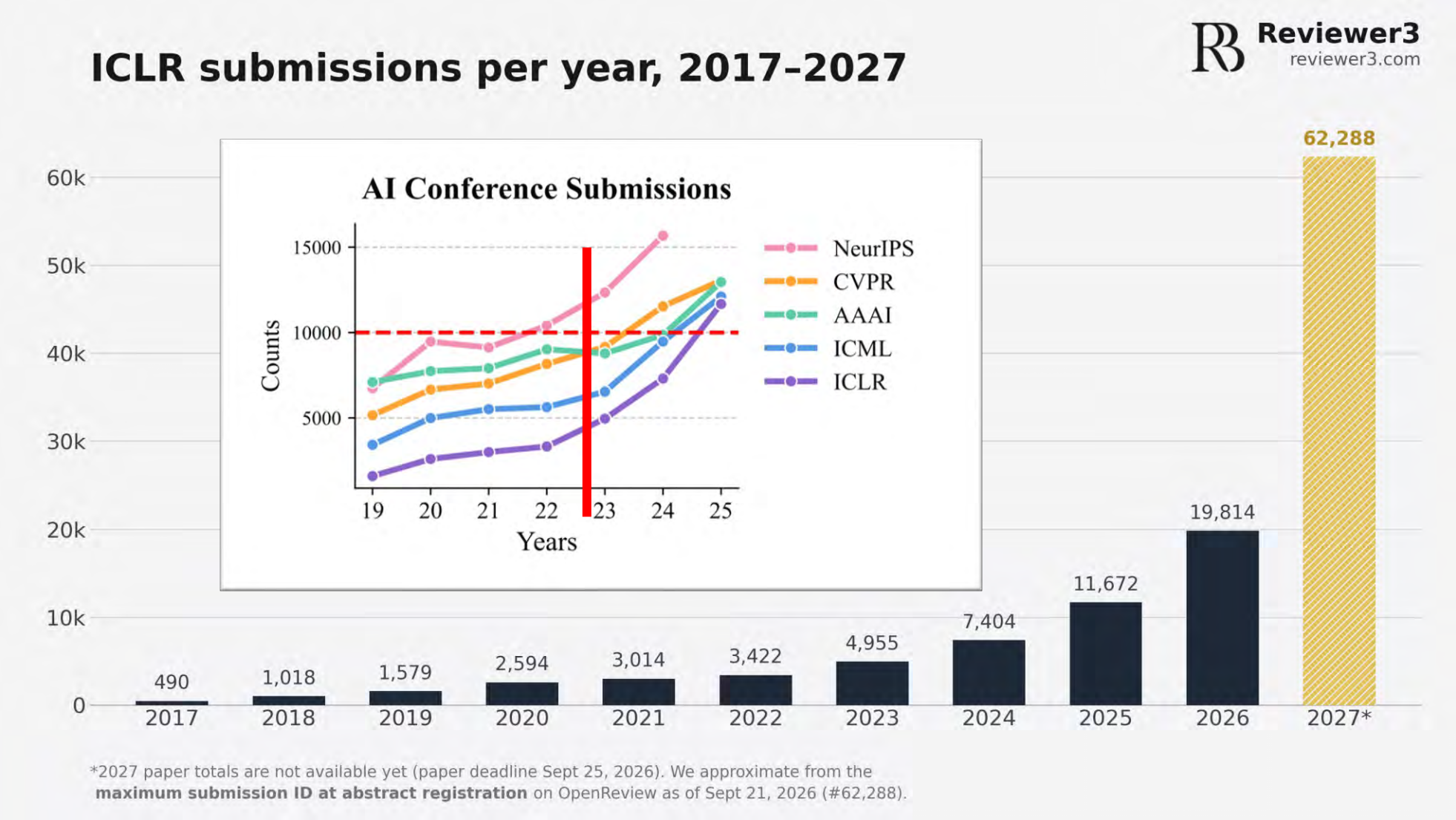
Example (2026)	Challenge / why it matters	What the AI did	Reported / est. AI cost	Human baseline / validation & status
Anthropic (Claude): ART enzyme system	Genome mining: discovering new biological systems among >200,000 reverse transcriptases.	~950 AI agents searched >200,000 reverse transcriptases in 21 hours and identified ART, a previously unrecognized repeat-associated system found mainly in bacteriophages; 210M tokens were used. Anthropic	~$2k–$10k*	Anthropic estimates comparable expert analysis would take weeks to months. Humans are running the wet-lab experiments; biological function remains unknown. Status: preprint / company report. (Anthropic)
OpenAI: Navier–Stokes	$1M Millennium Prize Problem: a central problem in mathematics and fluid mechanics, unresolved for ~90 years.	~10,000 AI agents produced a proposed resolution in 88 hours, claiming finite-time blow-up under smooth forcing; another 17 hours produced a machine-checked Lean formalization. OpenAI	~$6.5M+*	Generations of mathematicians worked on the problem. Clay says it has “apparently been settled”; formal evaluation and attribution are pending, and the unforced case remains open. Status: Lean-formalized; Clay review pending. (Clay Mathematics Institute)
OpenAI: 10 mathematics results	Ten long-standing open problems spanning geometry, coding theory, group theory, complexity, cryptography and combinatorics.	AI produced ten results that resolve or substantially advance open problems, including a proof of Ehrhart’s volume conjecture and a construction establishing that non-sofic groups exist; all ten arguments were then formalized in Lean. OpenAI	~$2k solution-search inference	The problems had resisted specialists for years to decades. Humans prepared the manuscripts with model assistance; all ten have Lean certificates, while broader independent mathematical review remains incomplete. Status: OpenAI release + Lean certificates. (OpenAI)
Anthropic (Claude): Fermat formalization	Formal verification: translating one of modern mathematics’ largest proof structures into a form a computer can verify line by line.	Claude completed the first end-to-end computer-checked formalization of Fermat’s Last Theorem in 11 days, producing ~13M lines of Lean and proofs of ~30,300 theorems. Anthropic	~$300k*	A community formalization begun in 2024 was expected to take years; Claude built on substantial existing human infrastructure. Lean verified the result using only its standard axioms, and a comparator confirmed that the statement matches Mathlib’s FLT statement. Status: machine-checked formal proof. (Anthropic)
FutureHouse (Robin): dry-AMD drug candidate	Drug discovery: turning hundreds of papers and experimental data into testable treatment hypotheses.	Robin synthesized 551 papers, proposed experiments and drugs, and identified ripasudil; its analysis found a 1.89× increase in RPE phagocytosis versus 1.75× in the human re-analysis. Nature	$10.76 per research run	The authors estimate 359–424 human cognitive hours versus <2 hours with Robin; humans selected candidates and performed the physical experiments. Status: peer-reviewed Nature paper + wet-lab validation. (Nature)
Google: AI Co-Scientist	Biomedical hypothesis discovery: generating mechanisms and treatment hypotheses that can survive experimental testing.	In 2 days, its top hypothesis matched the Imperial team’s then-unpublished experimental result: cf-PICIs spread across bacterial species by interacting with diverse phage tails; other drug and target hypotheses were also validated experimentally. Nature	Not disclosed	The Imperial team had been investigating the cf-PICI question for over 10 years, and the AI was tested before their result was publicly available. Wet-lab validation still required scientists. Status: peer-reviewed Nature paper; experimentally validated. (Imperial College London)
University Hospital Cologne (SPARK): cancer pathology	Biomarker discovery: finding clinically relevant biological patterns across >5,400 patients, 18 cohorts and five cancer types.	SPARK generated 475 biological ideas, translated them into 2,368 analytical parameters, retained 1,115 non-redundant parameters, and found associations with prognosis, established pathology variables and predictive biomarkers. Nature Medicine	~$4k–$5k†	Conventional studies require researchers to invent and implement candidate pathology features manually; SPARK automated the idea-to-analysis loop at scale. Prospective clinical validation is still required. Status: peer-reviewed Nature Medicine paper. (Nature)
AutoRecLab (ours): RecSys experimentation	Autonomous RecSys experimentation: turning a natural-language research prompt into an executable empirical study.	AI derived experiment requirements, wrote and executed the code, generated plots and analyses, and in a random-seed study reproduced qualitative trends reported in prior human RecSys research. ACM RecSys ’26 paper	~$1 per run	An experiment of the kind AutoRecLab supports typically requires ~1–3 researcher-days‡ for implementation, debugging, execution and plotting; researchers remain responsible for experimental design and interpretation. Status: ACM RecSys ’26 Demo / early proof of concept.

* Estimated AI cost based on reported model usage; actual internal costs may differ. † SPARK reports total development/implementation cost, not pure inference cost. ‡ AutoRecLab human effort is a rough estimate, not measured in the paper.
General note: AI costs exclude model training, researcher salaries, laboratory work, equipment and later validation.

Consequences: Short term and long term



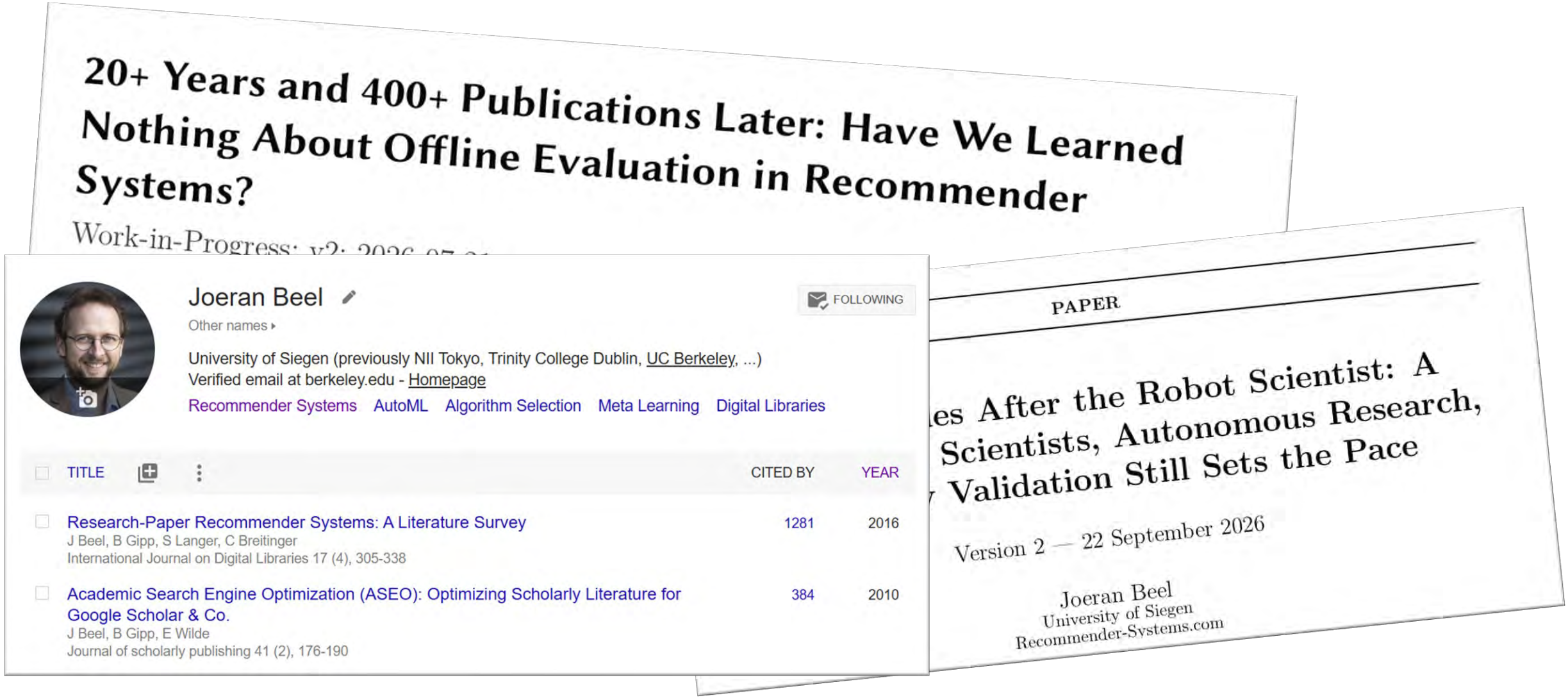
Submissions




Mentimet
Open Menti to edit
Replace this slide
Add Menti slides
menti.com
1268 6588

2. Tomorrow‘s Autonomous RecSys Research Labs

Surveys (with ChatGPT and Claude)



FRAME Workshop, Monday 13:30

A RecSys Paper for $20: Why We Must Build, Evaluate, and Govern Autonomous Recommender Systems Research Labs (AutoRecLabs)

Joeran Beel[1,2,*], Bela Gipp[3], Tobias Vente[4], Moritz Baumgart[1,3], Philipp Meister[3] and Sinan Pourazari[1]

[1]University of Siegen, Siegen, Gern
[2]Recommender-Systems.com, Düss
[3]University of Göttingen, Göttinge
[4]University of Antwerp, Antwerp,

Demo, Tuesday 17:30

AutoRecLab: An Autonomous Recommender Systems Lab

Moritz Baumgart
University of Siegen
Siegen, Germany
moritz.baumgart@uni-siegen.de

Philipp Meister
University of Siegen
Siegen, Germany
philipp.meister@uni-siegen.de

Justus Krell
University of Siegen
Siegen, Germany
justus.krell@student.uni-siegen.de

Michael Schmidt
University of Siegen
Siegen, Germany
michael3.schmidt@student.uni-siegen.de

Bela Gipp
University of Göttingen
Göttingen, Germany
bela.gipp@uni-goettingen.de

Joeran Beel
University of Siegen
Siegen, Germany
joeran.beel@uni-siegen.de

Abstract

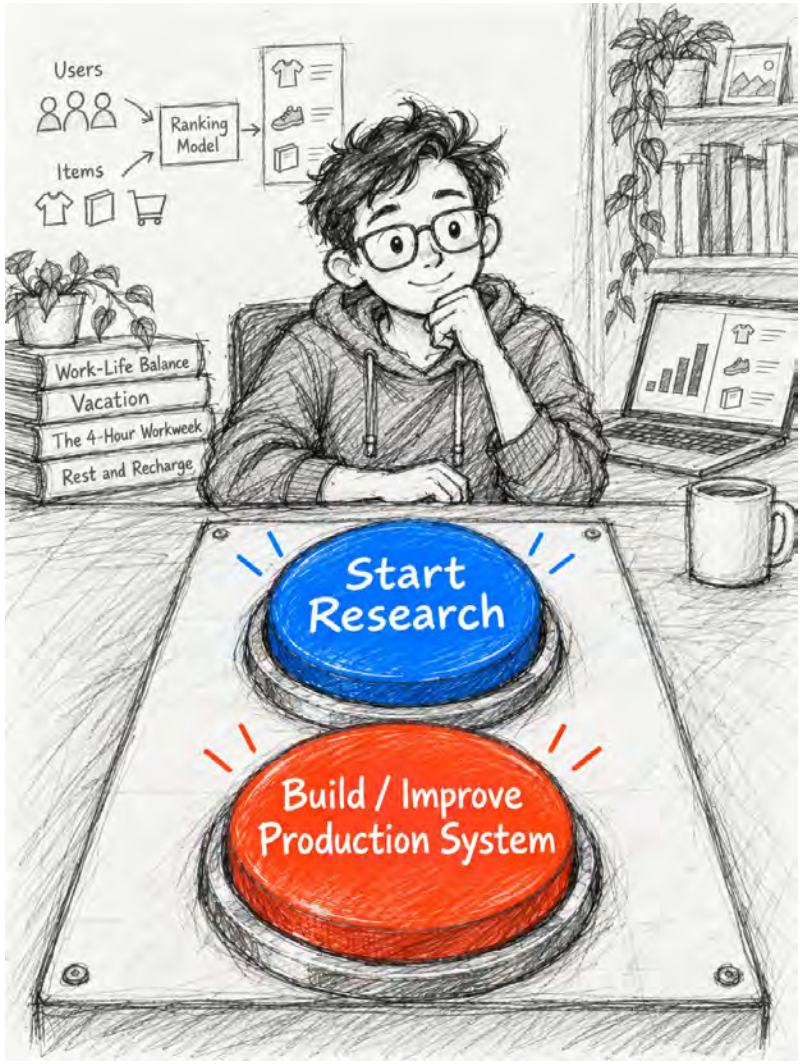
Recommender systems (RecSys) research depends on extensive empirical evaluation, yet translating experimental designs into executable code remains a manual, error-prone process. This paper

systems (RecSys) research is a promising domain for this automation, given that setting up and implementing experiments is inherently time- and labor-intensive. In particular, this includes writing custom data preprocessing, configuring evaluation loops, and adapt-

Today

- **Full/agentic research**
  - AutoRecLab
- **Production research**
  - Google Self-Evolving RS
  - Meta Auto-RecSys
  - AgenticRecTune
- **Agentic partial research**
  - Self-EvolveRec
- **AutoRecSys/HPO**
  - Auto-Surprise
  - LensKit-Auto
  - LibRec-Auto
  - AutoRec

Tomorrow



3. The Future of today‘s PhD students

Get used to be the boss (of AI Agents)

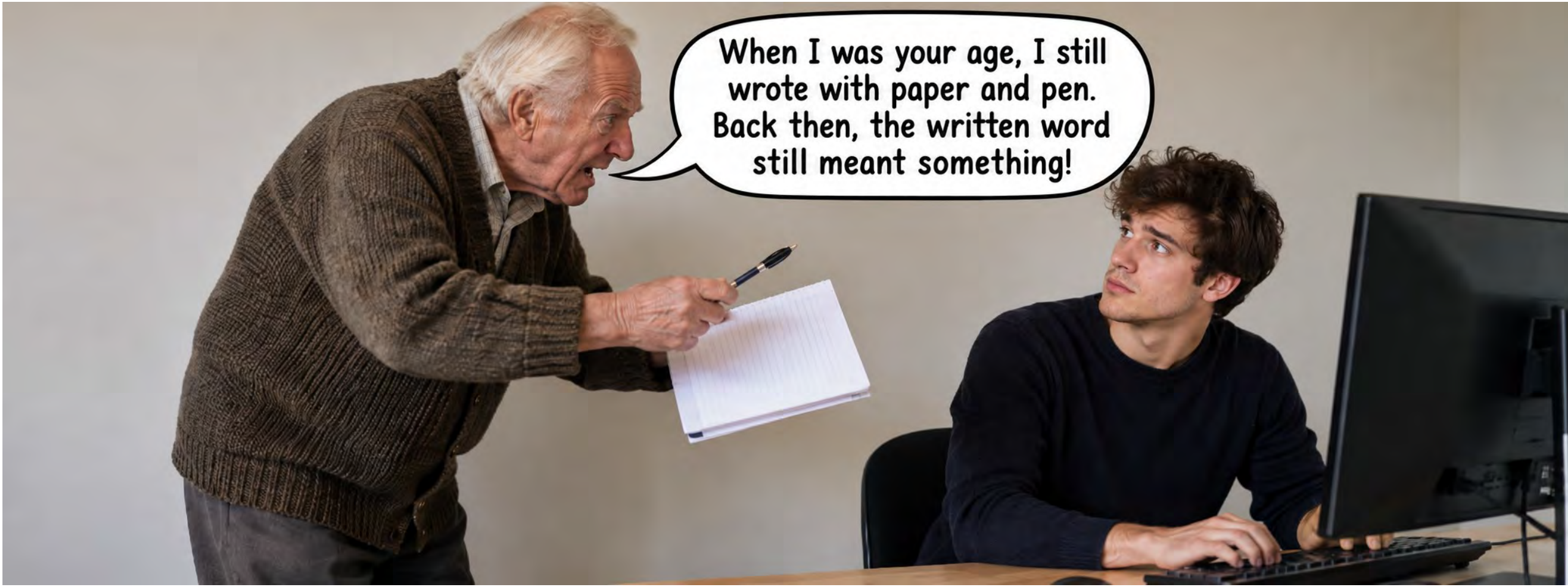
Embrace Change

Do more, do it really well, and sell it really well!


Minneapolis, MN, USA
27 September-2 October 2026
Association for Computing Machinery
Advancing Computing as a Science & Profession
1,307 distinct authors
1,510 author appearances
1,159 authors appear on exactly one paper
one author is co-author on seven papers
one paper with 28 authors
11 papers are single-authored
Industry papers are 32.8% of the proceedings, but account for 51.1% of all author appearances
Pinterest: 8 papers, 93 author appearances
Microsoft: 5 papers, exactly 5 Microsoft authors
149 of 250 titles contain a colon
83 titles start with a one-word method/system name followed by a colon
average title length: 10.1 words
longest title: 20 words, 136 characters
32 titles mention LLMs
17 mention agents/ agentic systems
"user" appears in twice as many titles as "item"
"generative" is the most common non-generic title word
the coauthorship graph has more than 160 separate clusters; the largest contains only 56 people
RecSys'26
Proceedings of the Twentieth ACM Conference on Recommender Systems
Gediminas Adomavicius, George Karypis and Joseph A. Konstan
General Co-Chairs
Bart Goethals, Martijn Willemsen and Minmin Chen
Program Co-Chairs
Arnab Bhadury, Felice Antonio Merra and Zhenhua Dong
Industry Co-Chairs

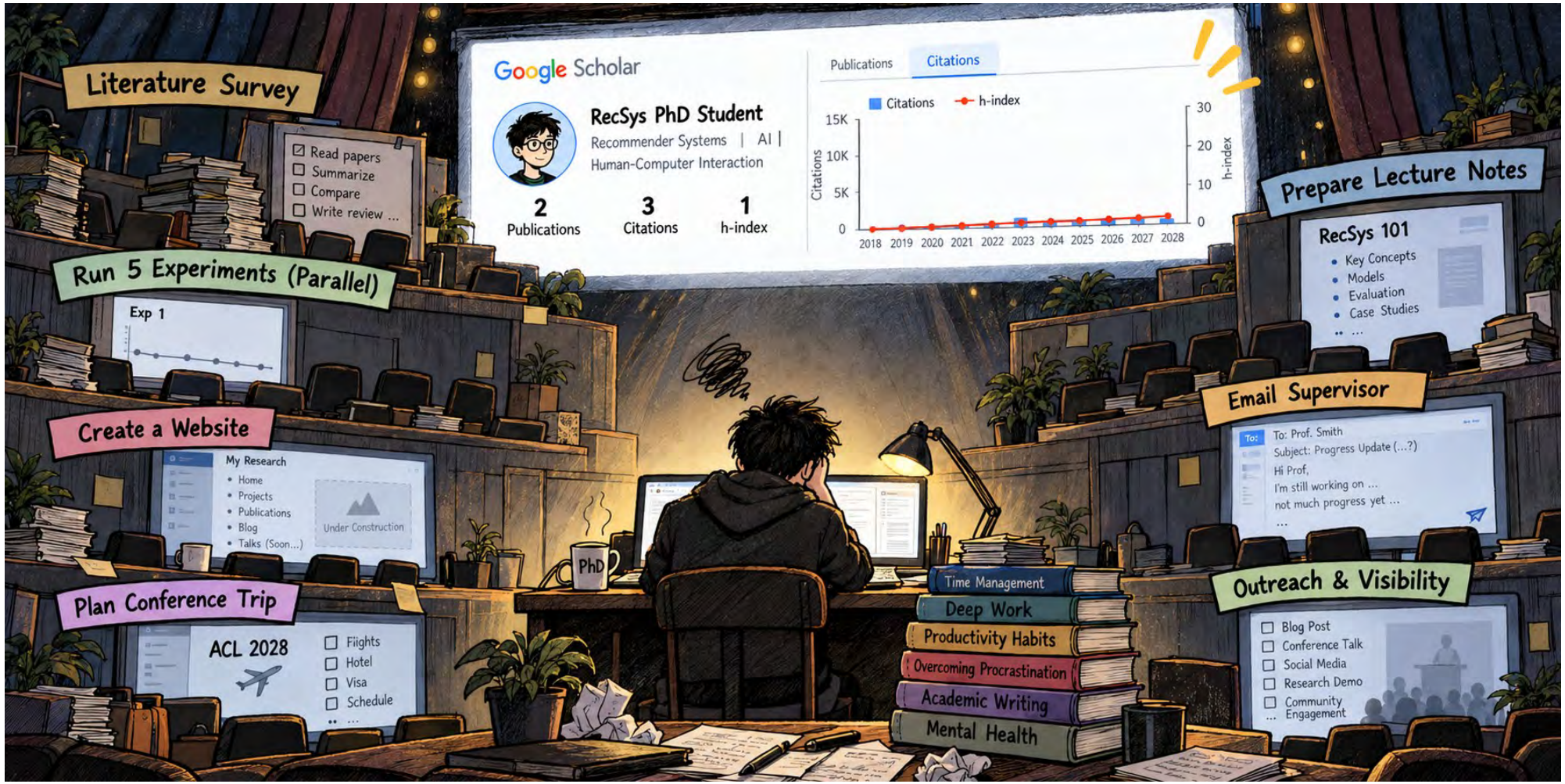
Be Critical and Brave (but don‘t jeopardize your PhD)

Realize, most of today‘s AI policies are made by old men who could be your grand fathers.



The future isn’t 10 work-hours per week. It’s 40+ or 0.

The “Alternative”



Prepare for Generation 4 of AI

- **Generation 3: Now**

  AutoResearch, needs human input, goals, resource allocation, prioritization, …

  → It needs people like you.

- **Generation 4:**

  Fully autonomous research, including new ideas, resource allocation, prioritization, …

  → Will you be needed? Who knows?

~~When~~ Because AI Does RecSys Experiments for $1: ~~Why~~ Spend Four Years on a PhD -- and truly achieve something!~~?~~

* There is one very important caveat though: today, all PhD students in academia are, more or less, equal. In the future, access to AI agents will make a massive difference on the impact and productivity of a PhD student.

Questions?

FRAME Workshop, Monday 13:30

beel.org

gscholar.beel.org

@jbeel

@JoeranBeel

Former Affiliations

Trinity College Dublin

NII National Institute of Informatics, Tokyo

A RecSys Paper for $20: Why We Must Build, Evaluate, and Govern Autonomous Recommender Systems Research Labs (AutoRecLabs)

Joeran Beel[1,2,*], Bela Gipp[3], Tobias Vente[4], Moritz Baumgart[1,3], Philipp Meister[3] and Sinan Pourazari[1]

[1]University of Siegen, Siegen, Gern
[2]Recommender-Systems.com, Düss
[3]University of Göttingen, Göttinge
[4]University of Antwerp, Antwerp,

Demo, Tuesday 17:30

AutoRecLab: An Autonomous Recommender Systems Lab

Moritz Baumgart
University of Siegen
Siegen, Germany
moritz.baumgart@uni-siegen.de

Philipp Meister
University of Siegen
Siegen, Germany
philipp.meister@uni-siegen.de

Justus Krell
University of Siegen
Siegen, Germany
justus.krell@student.uni-siegen.de

Michael Schmidt
University of Siegen
Siegen, Germany
michael3.schmidt@student.uni-siegen.de

Bela Gipp
University of Göttingen
Göttingen, Germany
bela.gipp@uni-goettingen.de

Joeran Beel
University of Siegen
Siegen, Germany
joeran.beel@uni-siegen.de

Abstract

Recommender systems (RecSys) research depends on extensive empirical evaluation, yet translating experimental designs into executable code remains a manual, error-prone process. This paper

systems (RecSys) research is a promising domain for this automation, given that setting up and implementing experiments is inherently time- and labor-intensive. In particular, this includes writing custom data preprocessing, configuring evaluation loops, and adapt-

for Artificial
mic Exchange
Service (DAAD)

Recommender-Systems.com
Founder