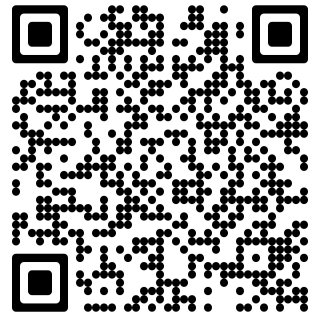




RESEARCH
ON RESEARCH
INSTITUTE

Funder Experiments, Experimental Funders

Tom Stafford, Senior Research Fellow
t.stafford@researchonresearch.org



tomstafford.github.io

TRANSFORMING RESEARCH

EXPERIMENTS
SYSTEMS
PLATFORMS

The Research on Research
Institute Impact Report
(2023-2025)

<https://researchonresearch.org/roris-impact-what-weve-achieved-and-where-were-going/>

Our definition of experiment

Principled: a research design that allows inference about what causes what (before/after, shadow experiments, true experiment/RCT)

Planned: primary outcome measure and analysis plan declared in advance

Public: a commitment to sharing the results regardless of outcome

AFIRE: Accelerator for Funder Experimentation

Forum

Sharing work by funders, for funders

Capacity building

Sprints on AI/ML in reviewer selection

Experiments

**Distributed Peer Review, Partial
Randomisation, Desk Rejection, and more!**

Experimental Funders Group, 2026

- 13 May 16:00 CEST- **AI in Funding**
- 17 June 16:00 CEST - **Distributed Peer Review**
- 14 October 16:00 CEST - **Desk Rejection**
- 16 December 09:00 CEST - **'Lottery First' methods**

Register!



<https://forms.gle/TV9DdsAHHP5aUDHN7>

Capacity building



The AFIRE AI in Funding Sprint brings together expertise on AI/ML, data science and experiment design across RoRI and RoRI partners.

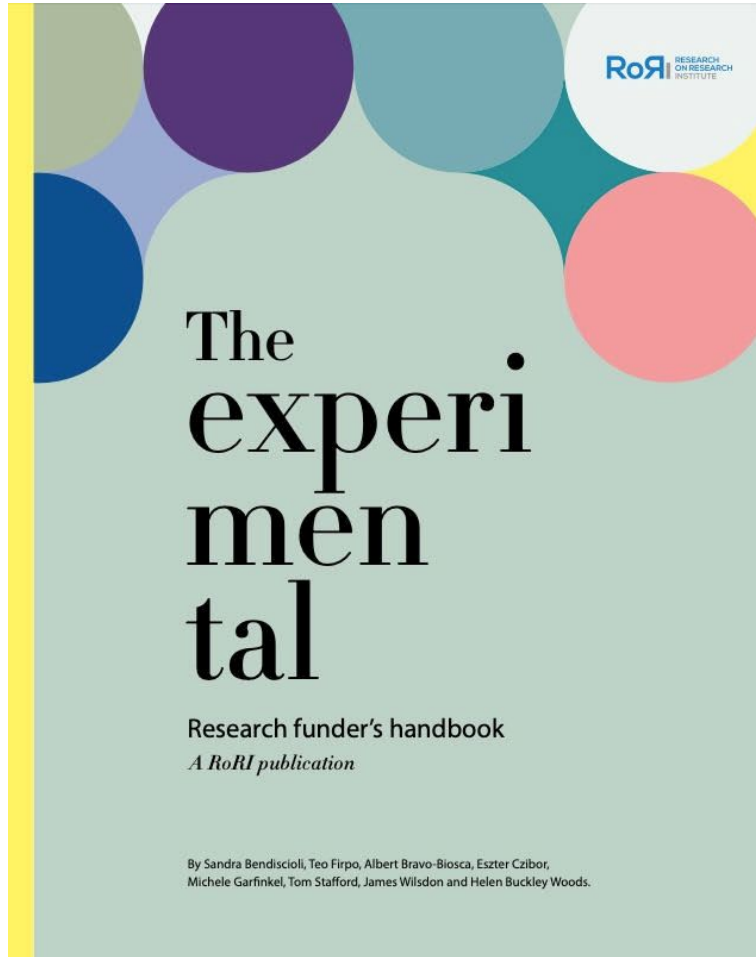
RoRI Summary note

Demand management: findings from a multi-funder workshop

Alyson Fox (Wellcome), Tom Stafford, Peter Kolarz (Research on Research Institute)

April 2026

We can plan/run experiments



AFIRE:

t.stafford@researchonresearch.org

josie.coburn@ucl.ac.uk

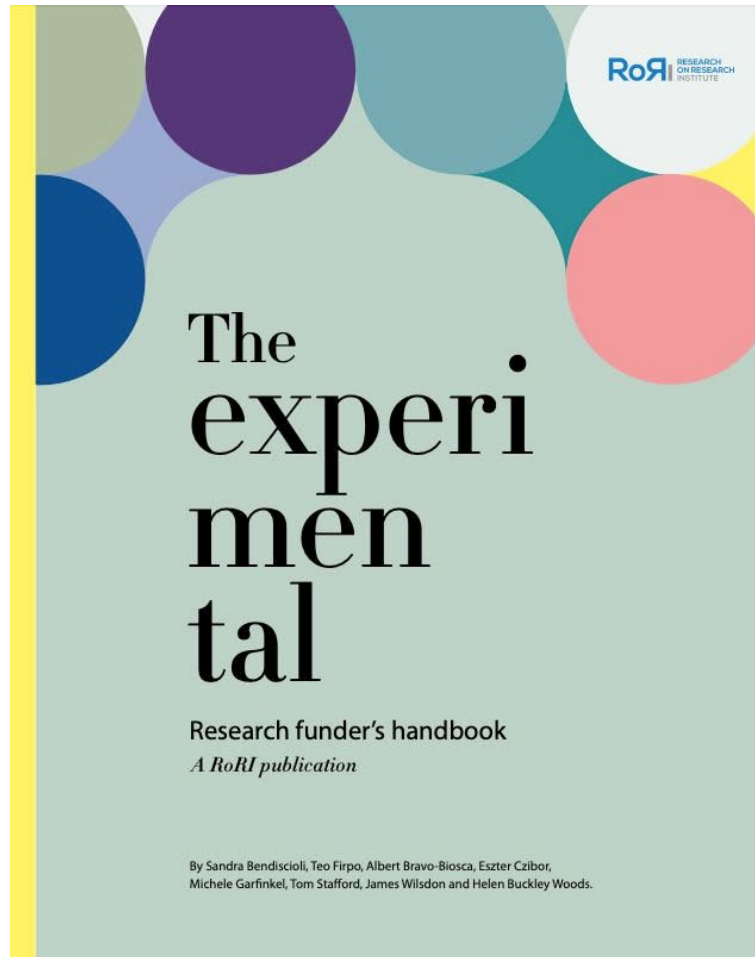
The experimental research funder's handbook (Revised edition, June 2022, ISBN 978-1-7397102-0-0).

<https://doi.org/10.6084/m9.figshare.19459328.v2>

These slides:

bit.ly/tomstafford

Experiments



AFIRE:

t.stafford@researchonresearch.org

josie.coburn@ucl.ac.uk

The experimental research funder's handbook (Revised edition, June 2022, ISBN 978-1-7397102-0-0).

<https://doi.org/10.6084/m9.figshare.19459328.v2>

These slides:

bit.ly/tomstafford

Partial Randomisation

Trials Catalogue

Funder	Dates
Health Research Council of New Zealand	2013-
VolkswagenStiftung	2017-2020
Austrian Science Fund (FWF)	2019-
Swiss National Science Foundation (SNSF)	2018-
Novo Nordisk Fonden	2022-2025
British Academy	2022-2025
UKRI / NERC	2022-
Wellcome	2023-
Nesta	2019-2020
University of Leeds	2023
UMC Utrecht/Ministry of OCW	2023

bit.ly/PRtrials



Bingo! Partial Randomisation for emergency research grants

How RoRI helped the Science Philanthropy Alliance with a funding conundrum



TOM STAFFORD
MAR 25, 2026



Share



Desk Rejection Shadow Experiment

Can agency staff predict those proposals with the least likelihood of success?

- a shadow experiment, not an intervention
- supports optimal use of external review
- UKRI leading participation
- recruiting schemes which will complete by end of 2026

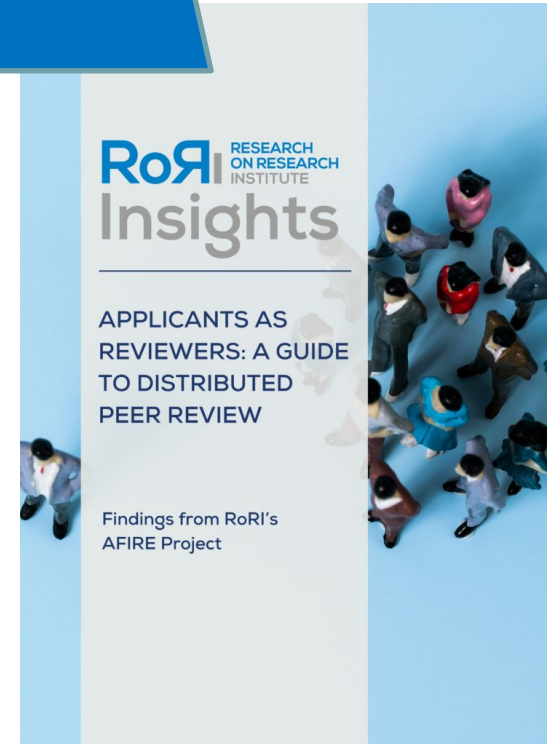
Enquiries:

Josie Coburn,
Research Fellow in Metascience,
Research on Research Institute
josie.coburn@ucl.ac.uk

Distributed Peer Review

Evaluating Distributed Peer Review at the Volkswagen Foundation

Anna Butters, Melanie Benson Marshall, Tom Stafford & Stephen Pinfield (Research on Research Institute and University of Sheffield);
Hanna Denecke, Alexander Bondarenko, Barbara Neubauer, Robert Nuske & Pierre Schwidlinski (Volkswagen Foundation)



AI reviewer matching@Metascience2025

Research Questions

1. Can language models help match proposals to reviewers?
2. Is it feasible for something like a conference to adopt/adapt this technology?
3. Can it be done securely/privacy respecting?



Maybe - evidence for meaningful improvements beyond human matching



Definitely yes



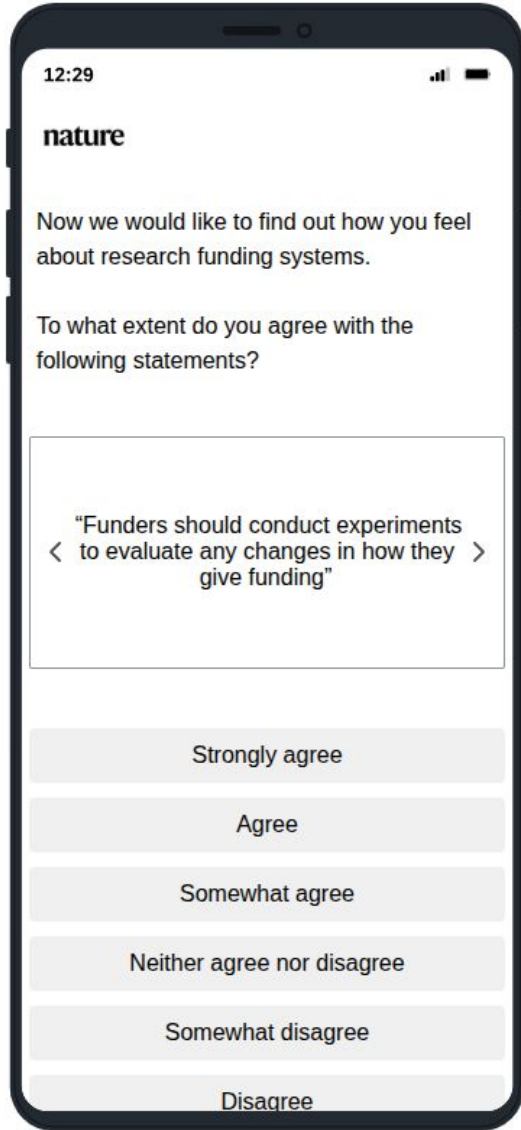
Definitely yes

Researcher Attitudes to Funder Expts

Where do researchers want innovation?

What should their motives be?

What makes experiments more or less acceptable?



12:29

nature

Now we would like to find out how you feel about research funding systems.

To what extent do you agree with the following statements?

“Funders should conduct experiments to evaluate any changes in how they give funding”

Strongly agree

Agree

Somewhat agree

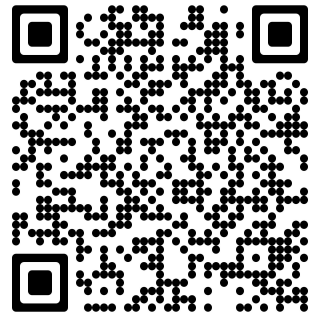
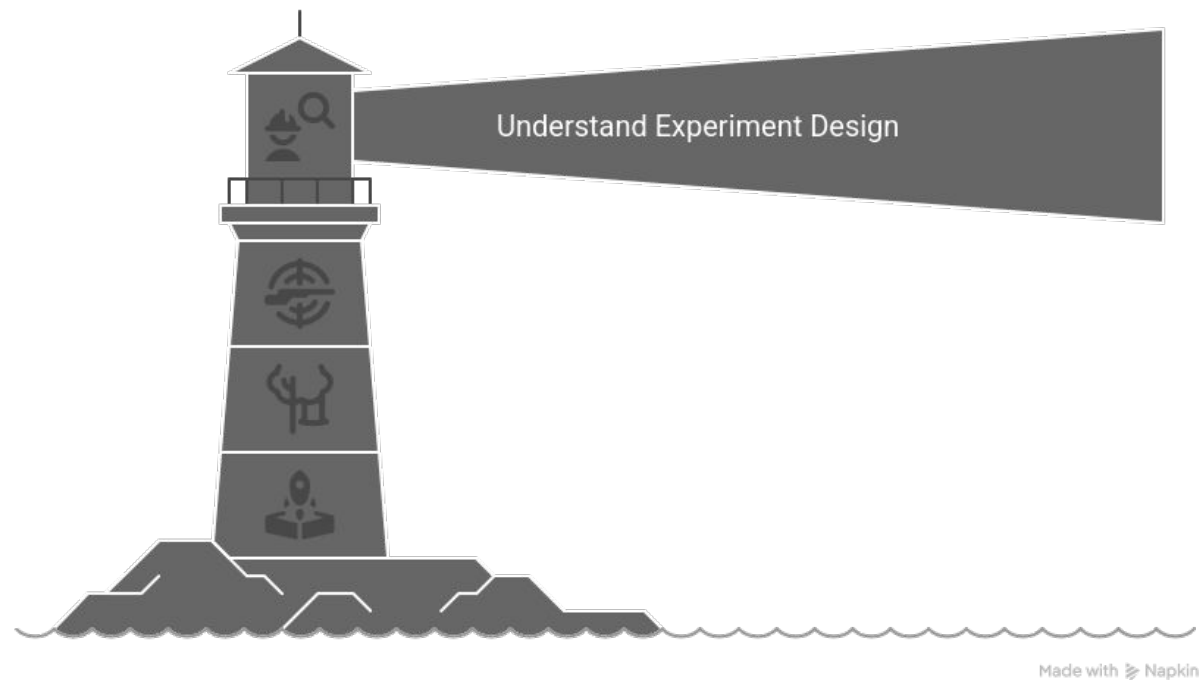
Neither agree nor disagree

Somewhat disagree

Disagree

SPRINGER NATURE

Becoming experimental



tomstafford.github.io

Get in touch!

researchonresearch.org
@RoRIInstitute



t.stafford@researchonresearch.org



tomstafford.github.io

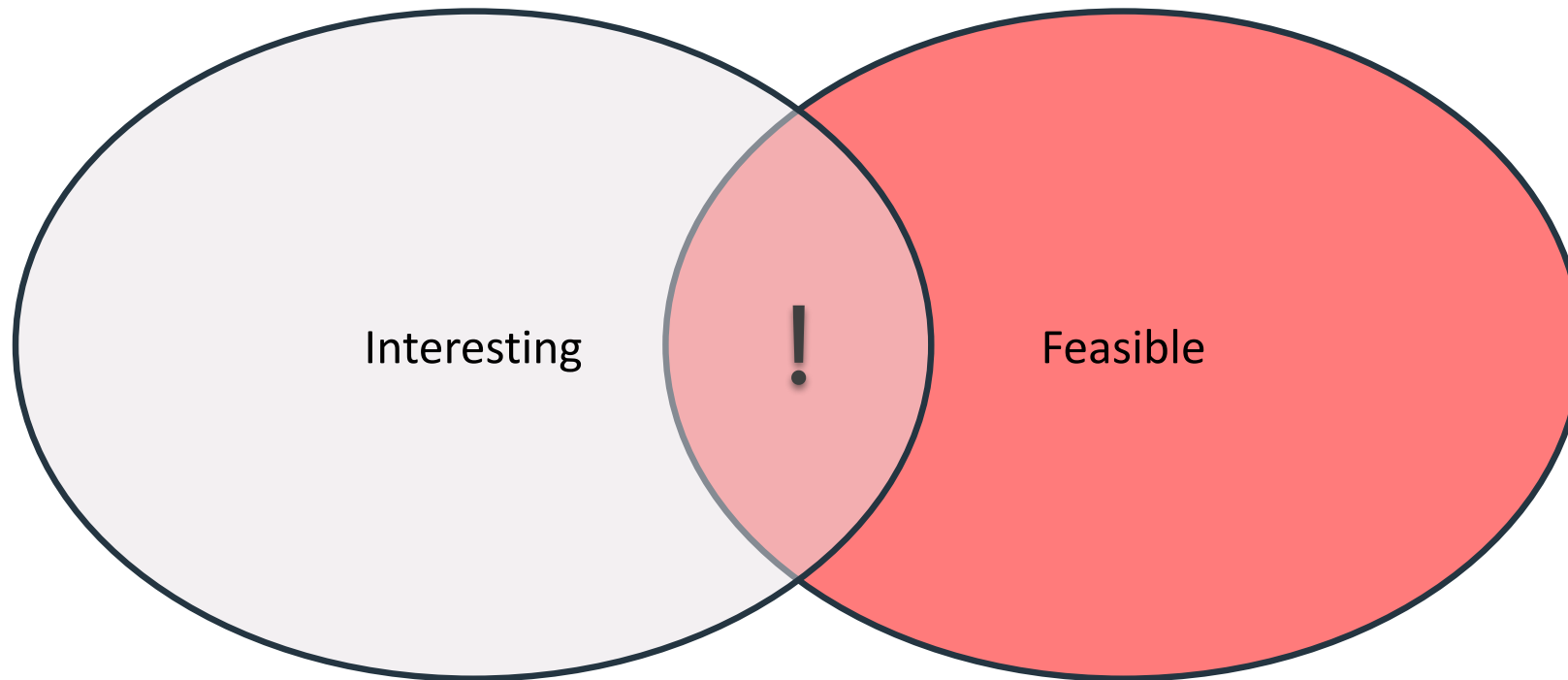
END

(reserve slides
follow)

Other AFIRE experiments

Thinking about experiments

“The Art of the soluble”



Good outcome measures



A part of the fresco "Triumph of Galatea," created by Raphael around 1512 for the Villa Farnesina in Rome. [Art Images via Getty Images](#)

Assays and microscopes

Yes/No

but is it the
right
question



Close view

but what are
you looking for?



Image: [CC Wikimedia](#)

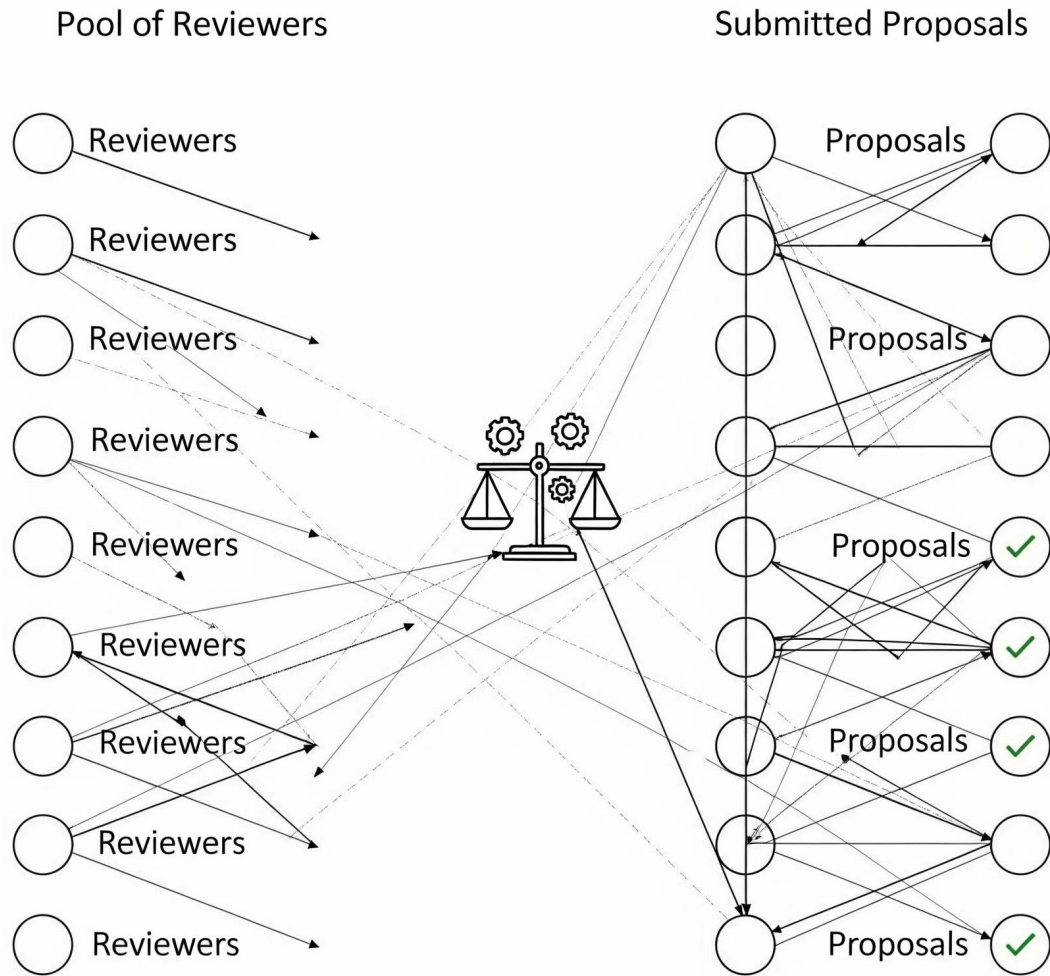
Just designing experiments is valuable



The Metascience 2025 conference experiment

Can AI be used for better matching of proposals to reviewers? Feasibility and formal evaluation with the Metascience 2025 conference

Josie Coburn and Tom Stafford 2025-11-26



Matching Algorithm
Assignment Process

Finding (enough, good) reviewers is a conceptual and practical problem

- conceptual: what makes a reviewer good?
- practical: how do you get a reviewer to agree?

Reviewer-proposal matching identified by GRAIL as a key area for possible experiments

Many funders already exploring this

Algorithms need validation!

Meta-metascience

AFIRE Commitment: Observation is not enough - we have to try things!

- demonstrate feasibility
- opportunity for better causal inference

Metascience 2025 conference, London

- a chance to show we'll take our own medicine
- appropriate domain for demonstrating feasibility
- added value: validate by collecting reviewer self-perception of suitability

The “shadow” experiment

Consent from those submitting and reviewers

All analyses done after final programme decisions

All analyses local - no data left the conference

441 submissions: Title, Abstracts

25 reviewers: assigned to submissions via keywords

1,323 reviews

- for each we have a match scores & a reviewer suitability judgement
- (each proposal seen by 3 reviewers)

Research Questions

1. Can language models help match proposals to reviewers?
2. Is it feasible for something like a conference to adopt/adapt this technology?
3. Can it be done securely/privacy respecting?

Matching - via embedding

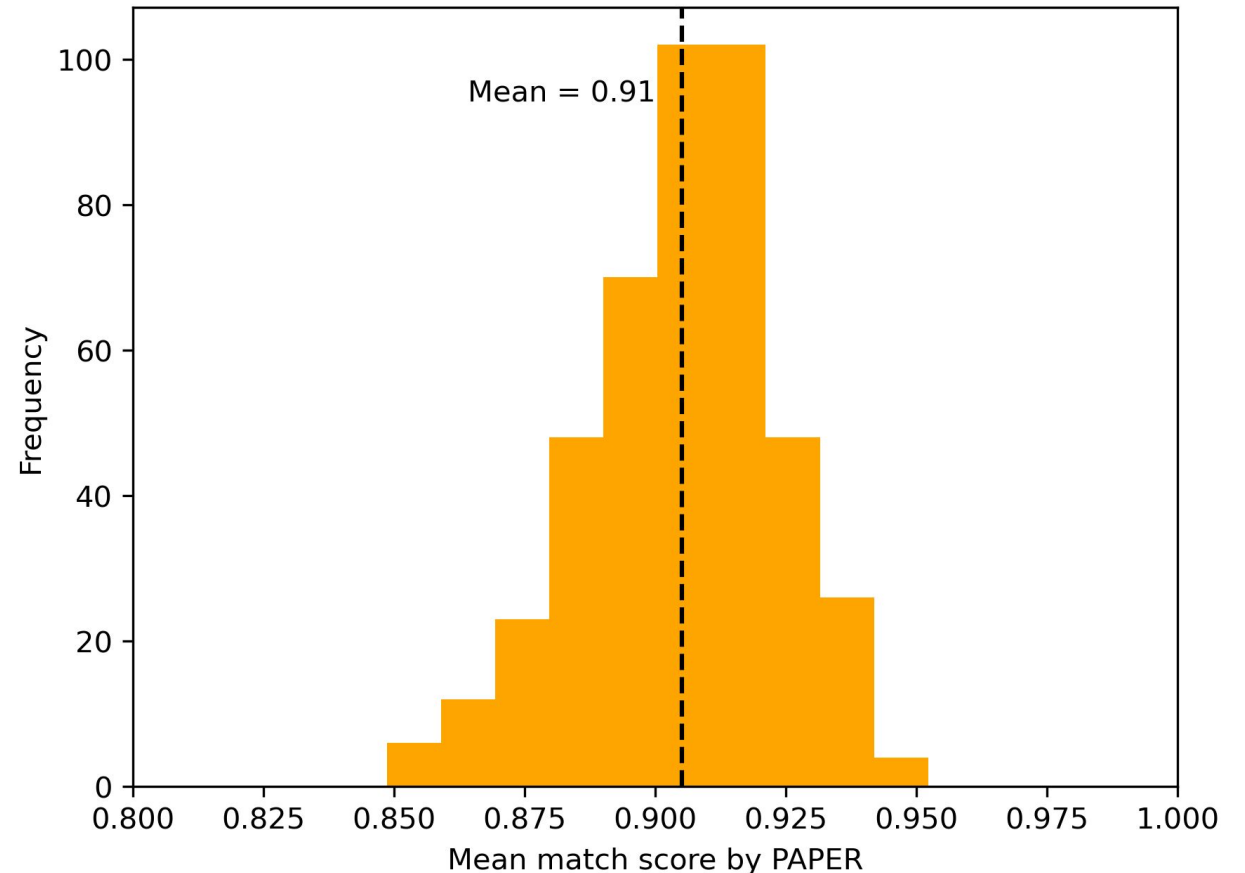
Reviewer keywords & proposal title+abstract -> embedding space

Code from SNSF:

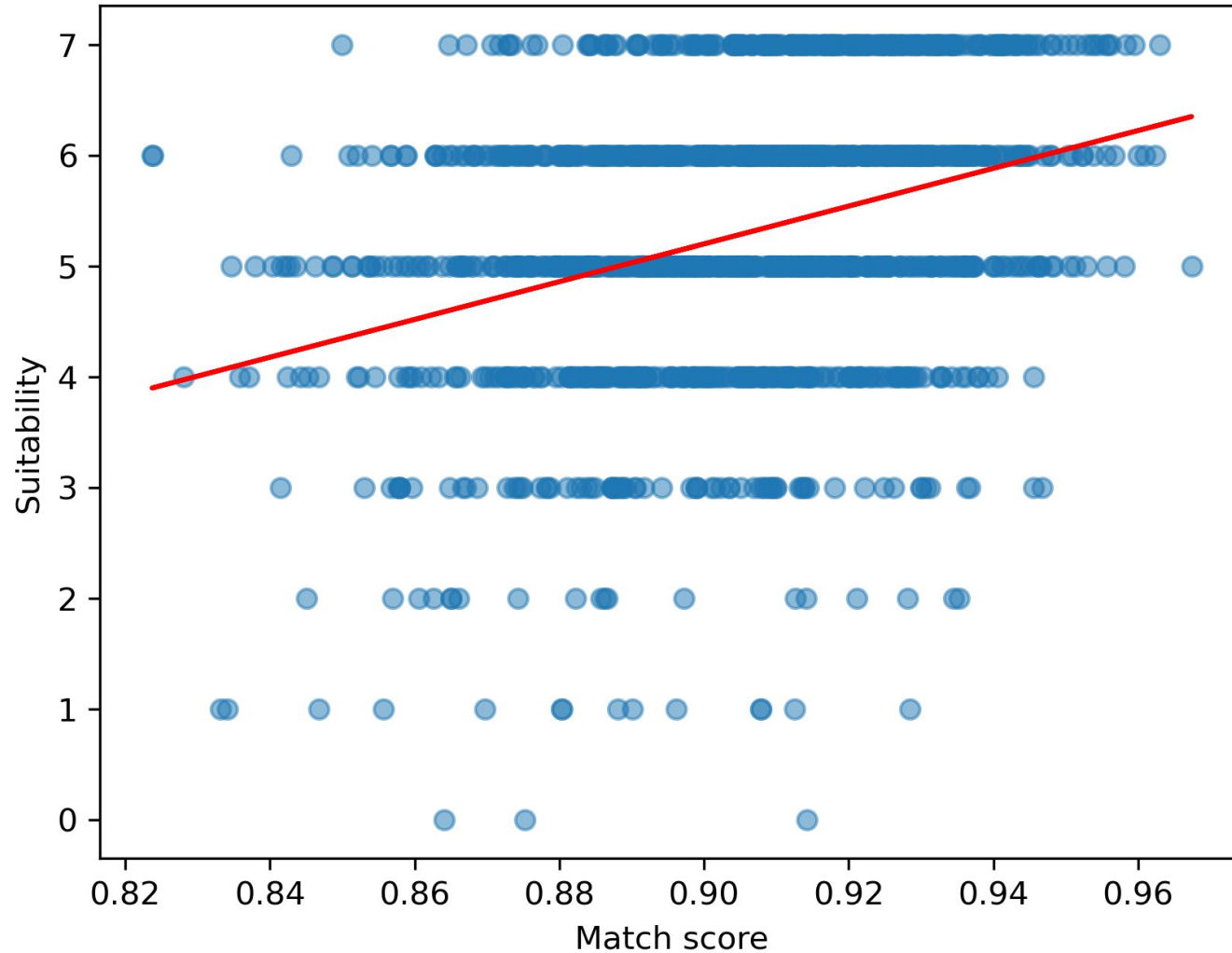
<https://github.com/snsf-data/snsf-grant-similarity>

- thanks to Gabriel Okasa and the SNSF data team!

Model: [SPECTER2: BERT model pre-trained on scientific texts and augmented by a citation graph](#)

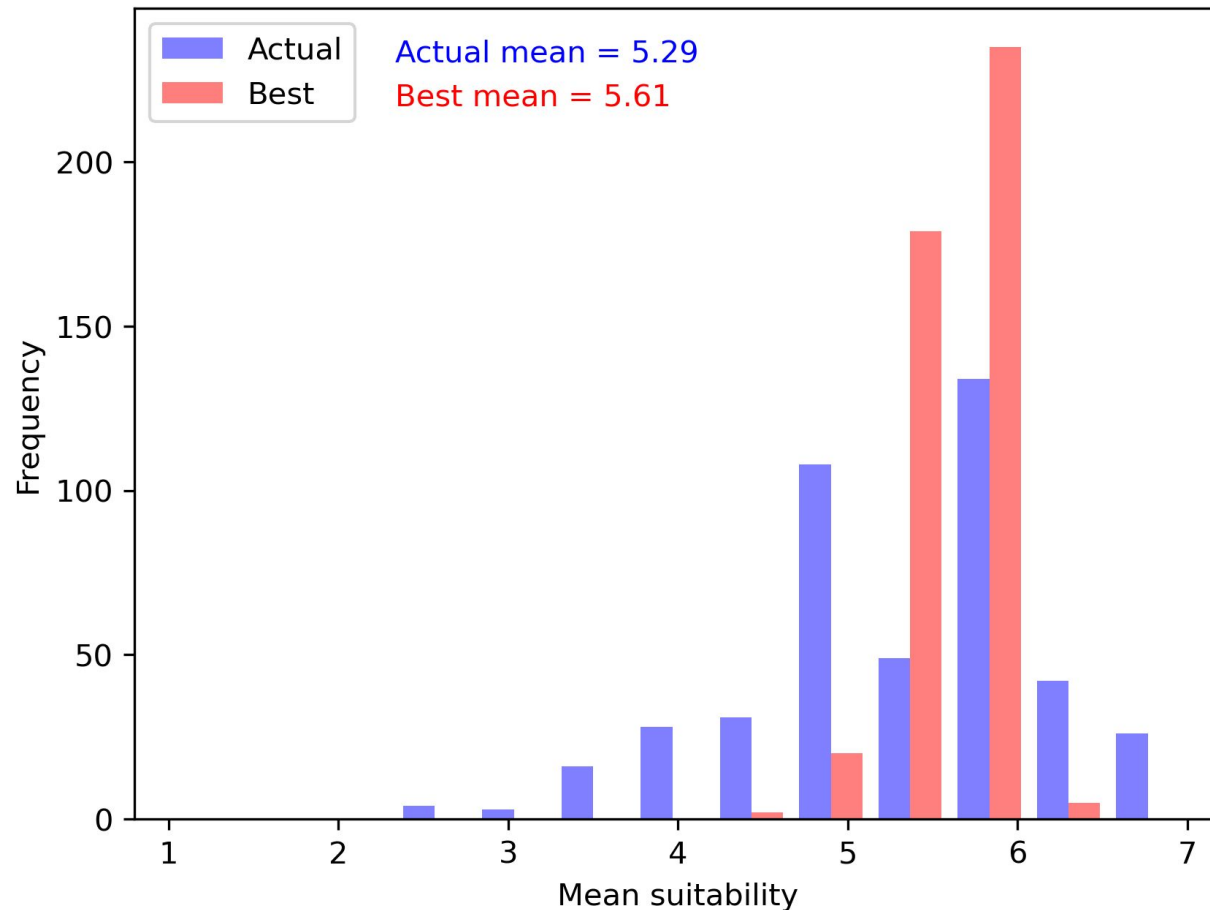


You can predict suitability from matchl



...and from this you can predict gain in suitability from using the optimal match

Actual and Predicted Best reviewer suitability ratings



Research Questions

1. Can language models help match proposals to reviewers?
2. Is it feasible for something like a conference to adopt/adapt this technology?
3. Can it be done securely/privacy respecting?



Maybe - evidence for meaningful improvements beyond human matching



Definitely yes



Definitely yes

**Thanks to all
participants!**