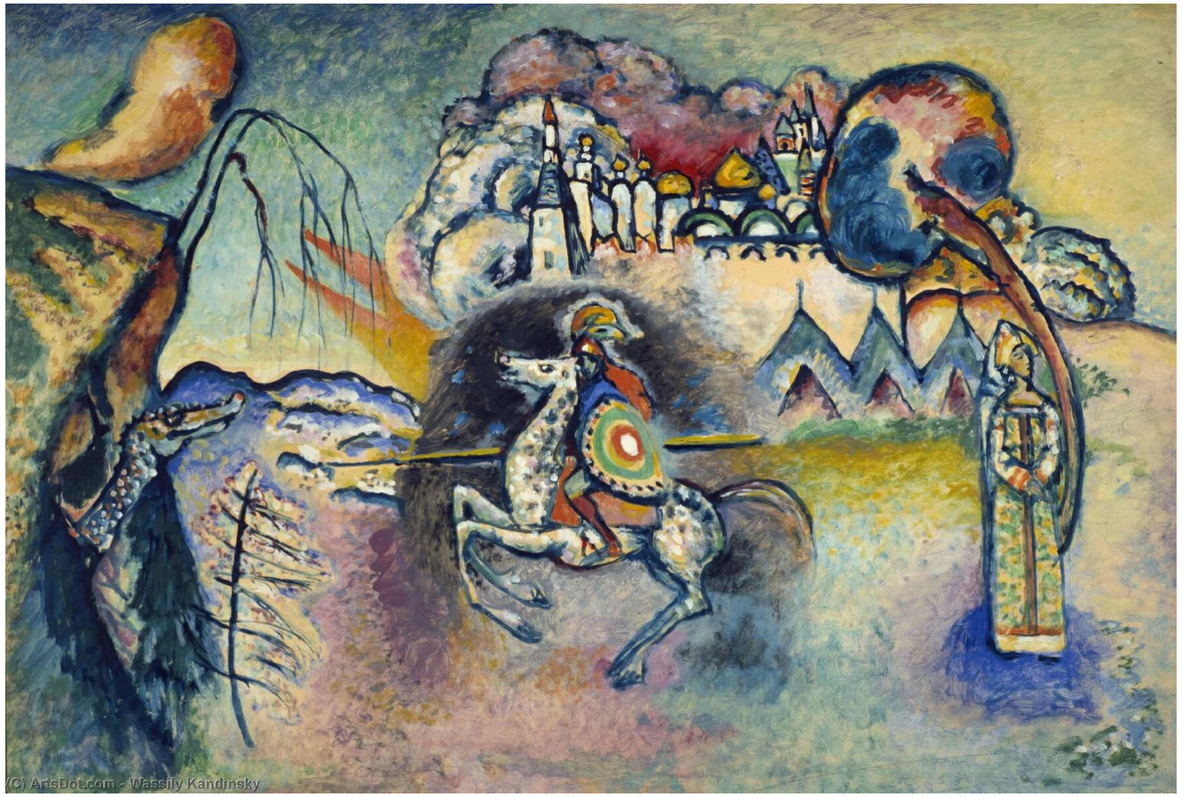
Evaluating Researchers in the AI Era



João Ribeiro

AI was used as an aid for searching information. No AI was used for preparing these slides.

A bit of context

Evaluation is a key component of an academic career:

- We're evaluated when applying for PhD positions;
- We're evaluated when applying for postdoc positions;
- We're evaluated when applying for tenure-track positions;
- We're evaluated during tenure review, and for subsequent promotions;
- We're evaluated when applying for funding;
- On top of this, in Portugal we're also evaluated every 3 years.

- We're mostly evaluated on our research;
- AI is dramatically changing the way we do research;
- This affects mathematical subjects more than other areas.

A bit of context

Evaluation is a key component of an academic career:

- We're evaluated when applying for PhD positions;
- **We're evaluated when applying for postdoc positions;**
- **We're evaluated when applying for tenure-track positions;**
- **We're evaluated during tenure review, and for subsequent promotions;**
- We're evaluated when applying for funding;
- **On top of this, in Portugal we're also separately evaluated every 3 years.**

- We're mostly evaluated on our research;
- AI is dramatically changing the way we do research;
- This affects mathematical subjects more than other areas.

What’s this?

A starting point for an open discussion about evaluation, broadly construed:

- What should change in the way we evaluate researchers (in hiring, promotions, 3-year evaluations…)?

- How should the department and its members act now to ensure that evaluation practices are updated the way we want them to?

What should change in faculty/postdoc hiring?

What are other institutions/countries doing?

- Focus on qualitative aspects:
  - Absolute best work instead of number of papers in good venues.
- In-depth knowledge testing:
  - Replace high-level job talks by deep dives on candidate's work, led by experts in the department.
- Statement on the use of AI in application materials and in prior work.
- Valuing community-building activities beyond scientific production (e.g., dissemination).

What should change in faculty/postdoc hiring?

Concrete example 1: UCLA CS (https://recruit.apo.ucla.edu/JPF11410)

In evaluating candidates, the department values a strong record of meaningful research accomplishments over the number of papers published at top conference venues. Applicants are encouraged to identify in their research statement up to three contributions they consider most significant and to explain their impact.

[…]

In addition to a research talk and one-on-one meetings, interviews include a one-hour "deep dive" session in which a small group of faculty experts asks specific questions about the candidate's work, giving the candidate an opportunity to show the depth of their understanding.

What should change in faculty/postdoc hiring?

Concrete example 2: MIT Math (https://math.mit.edu/about/employment)

We request that applications and other materials, including (a) curriculum vitae, (b) publication list (with hyperlinks, if possible), (c) research statement, (d) an optional teaching statement, (e) a broader impact statement, (f) a Gen AI statement, and (g) three letters of recommendation be submitted online at www.mathjobs.org

The Gen AI statement is a detailed factual statement on the use of generative AI, if any, in the applicant's past work and current application materials.

What should change in faculty/postdoc hiring?

Concrete example 3: CNRS Math (https://cn.math.cnrs.fr/Doc/Advice_CR2027.pdf)

1 Uses of generative AI by the committee and candidates

The committee makes the following pledge:

- We will not put any part of any application in any generative AI tool.
- We will not use generative AI tools to make a decision on the quality of an application.
- We will not penalize the candidates either for using generative AI tools in their research or for not using them.

The committee asks the following from the candidates:

- To write their report on previous research and their research project without the assistance of generative AI tools.
- To include a declaration about the use of generative AI in their research, which explains, for each of the scientific productions highlighted in the application, what the methods, goals and results of that use were.

What should change in faculty/postdoc hiring?

Concrete example 3: CNRS Math (https://cn.math.cnrs.fr/Doc/Advice_CR2027.pdf)

2 Application materials

- Instead of a full publication list, a selection of a few [1] significant scientific productions (articles, preprints, software, databases etc). The committee asks candidates to give a brief rationale for their choices.

- A report on previous research. This report is typically 8–10 pages long. The committee asks candidates to choose a maximum of three research contributions that they will highlight and put into broader context. If the candidates have used generative AI, we ask them to explain the methods, goals and outcomes of this use.

Note. The committee is fully aware that generative AI has lately caused very fast, and occasionally chaotic, evolutions in some research areas. This can significantly and unexpectedly impact research programmes, but the committee considers that it can nevertheless be interesting to pursue them, provided this is scientifically motivated. Indeed, the rapid proliferation of texts whose results have not been verified by mathematicians and are not up to the mathematical community's standards does not on its own constitute a definitive answer to the corresponding problems.

What should change in faculty/postdoc hiring?

Concrete example 3: CNRS Math (https://cn.math.cnrs.fr/Doc/Advice_CR2027.pdf)

- **Dissemination.** The committee will also pay attention to scientific outreach activities. In the present context and with the large-scale deployment of generative AI tools, transmitting scientific knowledge and putting it into perspective becomes even more important.

5 Interviews

The interview format will evolve to allow more in-depth discussion with the committee about the candidates' previous research and research programme. For the 2027 interviews, we are planning to implement the following format:

- Interviews consisting of a 10-minute presentation by the interviewee followed by about 30 minutes of discussion with the interview sub-committee.[3]
- Interview sub-committees composed of 9 members of the committee.[4]

What should change in faculty/postdoc hiring?

Many other examples discussed at https://proofsandprompts.com/2026/11/14/hiring-this-year-advice-and-changes-from-committees/:

- Postdoc hiring in Cambridge, ETH Zurich, MPIM;
- Applications for IAS postdocs/membership.

Other useful info at https://cmsa.fas.harvard.edu/media/2026/10/Summit-on-Math-Research-in-the-Age-of-AI.pdf

This isn’t just an IST problem

Hiring practices must change across all Portuguese universities.

Ideally, we would see a coordinated effort by the Portuguese mathematical community.

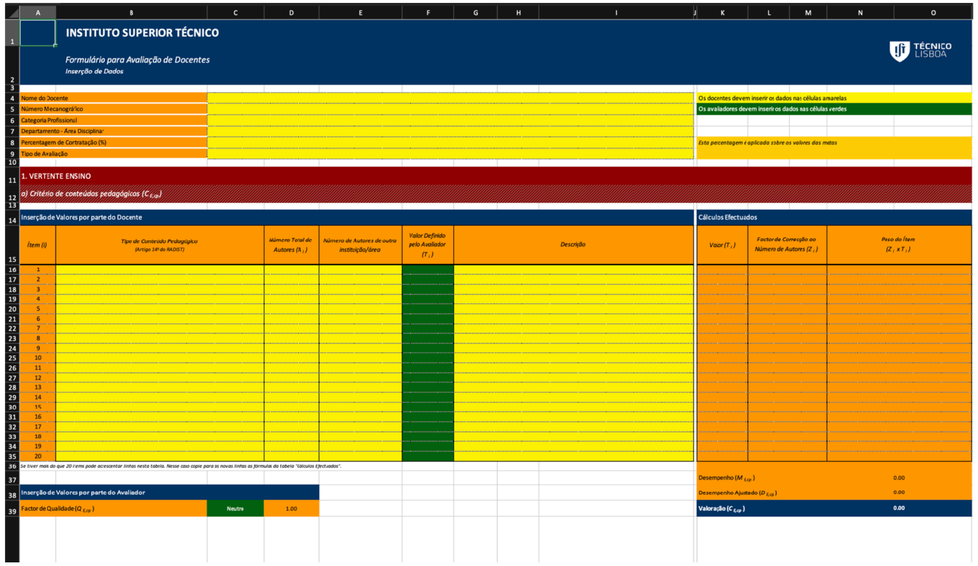
Example: The coordinated efforts of the US theory of computation community.

- Sept 2026: Convened a working group at the Simons Institute in Berkeley.
  - Released a report with recommendations (https://tinyurl.com/aitcs-simons)
  - Led to a statement on hiring in theory of computation signed by 100+ faculty (https://njbergam.github.io/tcs-hiring-statement/)
- Oct 2026+: Local advocates leverage this statement to change hiring practices at US institutions.
  - Recommendations already adopted by UCLA CS (discussed before).

Regular evaluations

Portuguese academics must be evaluated every 3 years by law.

Portuguese universities have, unfortunately, decided to implement this (and other evaluation procedures) via numerology.



INSTITUTO SUPERIOR TÉCNICO

Formulário para Avaliação de Docentes

Inserção de Dados

TÉCNICO LISBOA

Nome do Docente

Numero Mecanográfico

Categoria Profissional

Departamento - Área Disciplinar

Percentagem de Contratação (%)

Tipo de Avaliação

Os docentes devem inserir os dados nas células amarelas

Os avaliadores devem inserir os dados nas células verdes

Esta percentagem é aplicada sobre os valores dos mapas

1. VERTENTE ENSINO

a) Critério de conteúdos pedagógicos ($C_{1.a}$)

Inserção de Valores por parte do Docente

Item (i)	Tipo de Conteúdo Pedagógico (artigo 14º do RJUE)	Número Total de Autores (A_i)	Número de Autores de outra instituição/área	Valor Definido pelo Avaliador (T_i)	Descrição
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					

Se tiver mais do que 20 items pode acrescentar linhas nesta tabela. Nesse caso copie para as novas linhas as fórmulas da tabela "Cálculos Efectuados".

Inserção de Valores por parte do Avaliador

Factor de Qualidade ($Q_{1.a}$) — Neutro — 1.00

Cálculos Efectuados

Valor (T_i)	Factor de Correcção ao Número de Autores (Z_i)	Peso do Item ($Z_i \times T_i$)

Desempenho ($M_{1.a}$) 0.00

Desempenho Ajustado ($D_{1.a}$) 0.00

Valoração ($C_{1.a}$) 0.00

$$Q = \sum_{i=1}^{N} \frac{H_i}{A_i} \left(F_i + \frac{C_i}{5} \right)$$

(CGD/ULisboa prize scoring rules)

These practices were already anti-scientific before AI.

The new AI status quo gives us a great opportunity to correct this.

Regular evaluations

Just like hiring must become much more qualitative, so must these evaluations.

Example: CNRS

CNRS researchers are also evaluated periodically, by peers and in qualitative terms.

> The evaluation of CNRS researchers by the sections and interdisciplinary commissions of the National Committee is founded on four principles:
>
> 1. Quality: The evaluation only looks at the result and not at how prestigious the publication medium is;
> 2. Scope: The personal contribution made to output submitted for evaluation and the impact is considered more important than an exhaustive list;

Regular evaluations

Evaluation should be qualitative and should encourage slow but high-quality science.

- We should leverage AI to produce better work, not *more* work. Focus evaluation on a few selected contributions *(not necessarily papers)*.

- Place more value on community-building activities:
  - Expository work (new expositions of existing results, surveys…);
  - Outreach;
  - Mentoring;
  - Academic service (reviewing, conf committee membership, journal editorial boards…);

Summing up

- AI is changing the way we do research, and so the way we evaluate researchers must change too;
- **Postdoc/faculty hiring practices are already changing abroad:**
  - Focus on top x works + deep-dives + valuing of community-building activities;
  - Hiring practices in IST/Portugal must adapt quickly too. How should we go about it?
- **Coordinated community-wide efforts are important enablers of change:**
  - US TComp community: Simons workshop + letter + local lobbying at universities;
  - We should also coordinate efforts across Portuguese universities;
- **We now have a great opportunity to move away from numerology in Portuguese academia.**
  - How should we go about championing more qualitative and diverse evaluations?
- **It is extremely important that we take the initiative and come up with concrete proposals. Otherwise, someone else will do it for us, and it may damage mathematics in Portugal.**