



TECHNISCHE  
UNIVERSITÄT  
WIEN



# Evolving Practical Reproducibility in HPC

Sascha Hunold, TU Wien, Austria (SC24 Reproducibility Initiative Chair)



## Reproducibility Initiative at SC24 / Main Updates

### ■ AD/AE Appendices

- authors need to state the relationship between **artifact - contribution – paper** element
- defined **8h reproducibility threshold** for reviewers

### ■ Expanded size of Reproducibility Committee

- 101 members (still not enough)

### ■ Double-open reviews (if needed)

- allows for authors to **provide access to specific machines** (Fugaku, Frontier, etc.)

Artifact ID	Contributions Supported	Reproduced Paper Elements
$A_1$	$C_2$	Figure 7a–c
$A_2$	$C_3$	Figure 8

# New: Artifact Evaluation Reports

- reviewers provide **reports** on **why badges have been awarded**
- published as supplemental material
- increases **transparency of badge assignments**
- increases **visibility of reviewers**
  - should get credit for their work
  - reviewers could remain anonymous if they wanted to

## Artifact Evaluation Report

Marek Felšöci (MF)  
Centre Inria de l'Université de Lorraine  
et Antenne de Strasbourg  
Strasbourg, France  
marek.felsoci@inria.fr

Hugo Taboada (HT)  
CEA, DAM, DIF – Arpajon, France  
Université Paris-Saclay  
CEA Laboratoire en Informatique  
Haute Performance pour le Calcul et la simulation  
Bruyères le Châtel, France  
hugo.taboada@cea.fr

### I. OVERVIEW OF REPRODUCTION OF ARTIFACTS

The following table provides an overview of each computational artifact's reproducibility status. Artifact IDs correspond to those in the AD/AE Appendices.

Artifact ID	Available	Functional	Replicated
$A_1$	•	•	•
Badge awarded	yes	yes	yes

### II. REPRODUCTION OF COMPUTATIONAL ARTIFACTS

#### A. Timeline

- The artifact evaluation was carried out by MF on July 28 and July 30, 2024.
- The artifact evaluation was carried out by HT on July 17 and July 18, 2024.

#### C. Details on Artifact Reproduction

- MF followed the instructions provided by the authors except for the software deployment. Indeed, for the aforementioned reasons, MF had to use a different experimental testbed. This also led to differences in versions of certain software components resulting in the unavailability to replicate all of the results. However, MF was able to replicate the major part of the results from the paper.
- For Artifact A1, the figure in Section III-B represents the results replicated by MF from Fig. 5 in the paper. MF have observed a certain differences compared to the results in the paper, but these do not seem to be significant. The speed-ups have the exact same tendency and magnitude as in the paper.
- For Artifact A1, the figures in Section III-C represents the results replicated by MF from Fig. 3 in the paper. MF was unable to replicate some of the results. This is likely due to the differences in the software environment.

## Statistics for SC24

Badge Type	Applied for	Awarded
No badge	20	N/A
Available	22	34
Functional	15	22
Replicated	<b>45</b>	<b>25</b>
Total Badges	187	153

$$45 \cdot 3 + 15 \cdot 2 + 22 = 181$$



# Challenges for Reproducibility Initiative

**Biggest Challenge** with AD/AE Authors

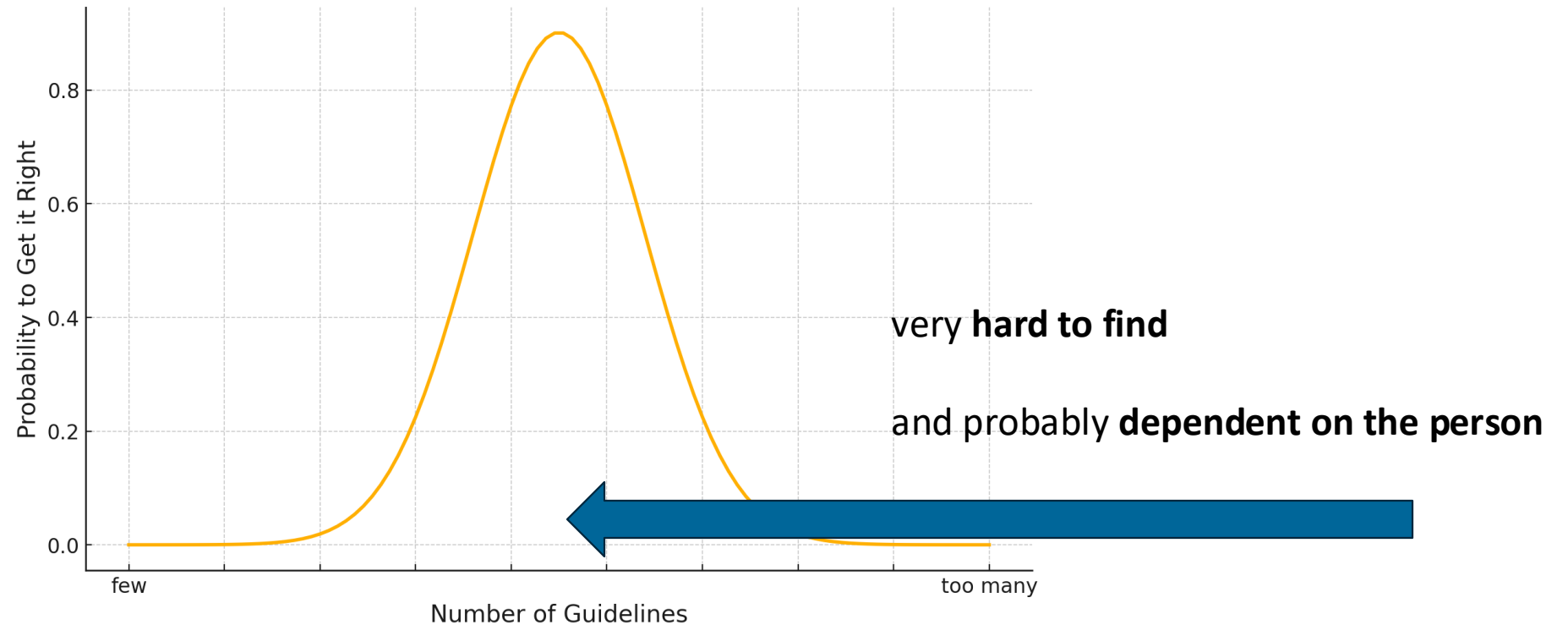
- authors do **not read the instructions** / guidelines

**Biggest Challenge** with AD/AE Reviewers

- reviewers do **not read the instructions** / guidelines

**Additional Challenge:** **Reproducibility Initiative is Not Reproducible**  
**would only receive badge: Artifact Available**

# Probability that ALL Authors or Reviewers get it Right



## Problems with Current Badges

### ■ Availability

- “Artifacts **do not need to have been formally evaluated** in order for an article to receive this badge. In addition, **they need not be complete** in the sense described above. “
- meaningless in the long run

### ■ Replicated

- For SC, artifacts can often only be replicated/reproduced with the authors’ help (on their machines).
- **too much burden on the reviewers**
- performance numbers are often impossible to reproduce: **cost of experiments outweigh purpose**

c



## Possible Solutions to Challenges in Reproducibility Initiative

- **all badges are meaningful** in different contexts
- for the **SC Reproducibility Initiative**:
  - **only check for Functionality!!!**
- functionality should be evaluated on an external system (not on authors' system)
  - **Chameleon Cloud** is the perfect experimental testbed for that
  - very important: **Artifact** should **only include the authors' contribution**
  - **no need to package competing algorithms/systems/libraries** etc. (was a huge problem for many authors)

