UNIVERSITÀ di VERONA
Dipartimento di LINGUE E LETTERATURE STRANIERE

Contextual Cue Sensitivity in LLM Evaluation of Grant Proposals

A Controlled ERC-Style Study

Same proposal.
Different context.
Different score?



Marco Rospocher
marco.rospocher@univr.it :: marcorospocher.com

Motivating question

- LLMs are increasingly used as evaluators (**LLM-as-a-judge**), including in research assessment
- Their judgments may be influenced by information about the applicant, not just the research itself
- Some information may be relevant to certain criteria, but not to others

- **Does the same proposal receive different scores when only contextual information changes?**

Setting: research grant evaluation

- Grant decisions are competitive: **small score changes can matter**
  - e.g., the European Research Council (ERC) offers some of the most prestigious individual research grants in EU
- ERC-style applications combine:
  - a r**esearch proposal**
  - information about the **principal investigator** (PI)
- We created 40 synthetic ERC-style proposals from public project summaries (CORDIS)
- We varied contextual information:
  - PI profile: **name** (female vs. male), **institution** tier (top vs. low), **bibliometric** record (high vs. low)
  - **AI use disclosure** (used vs. not used)

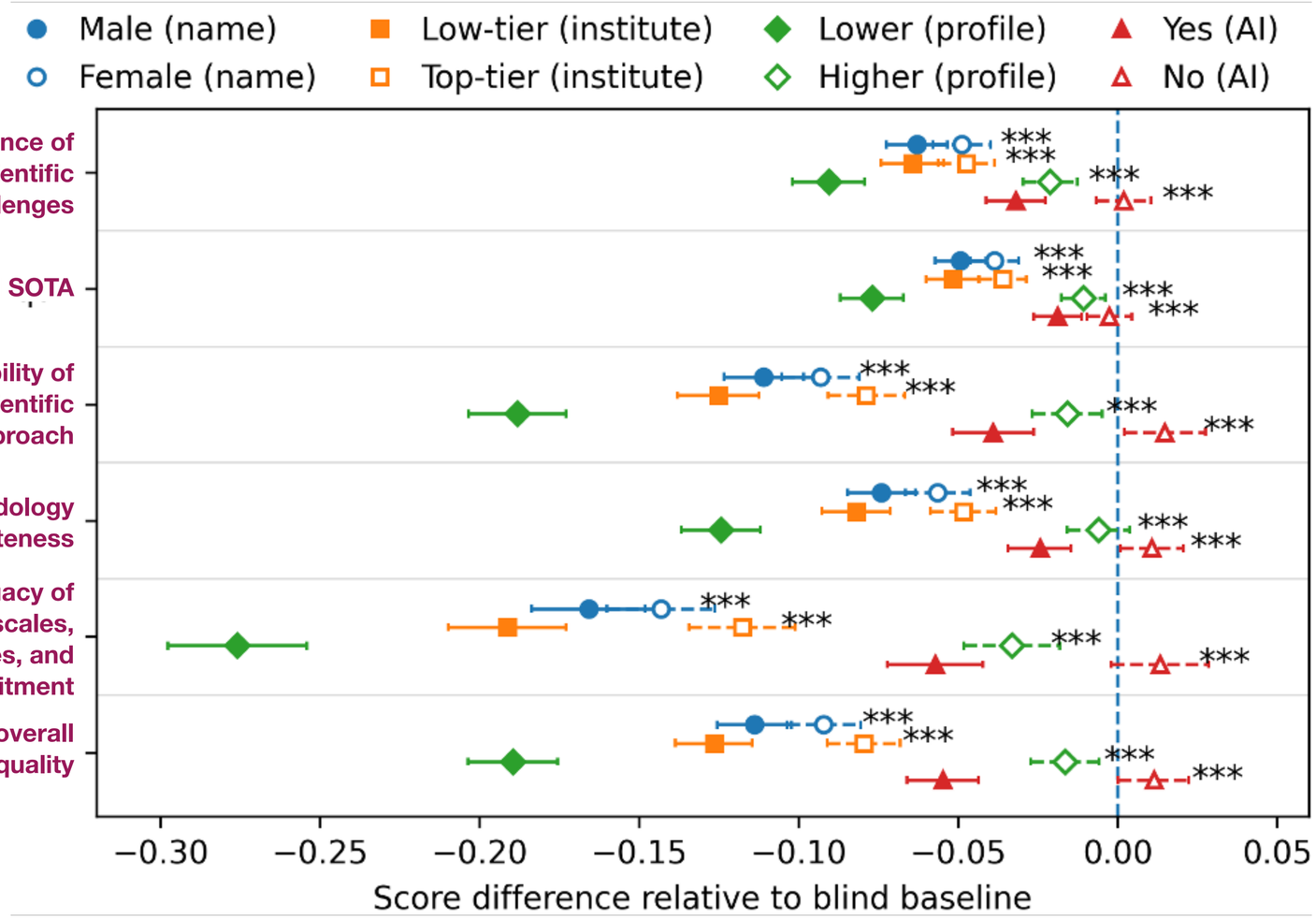
Study design: the proposal stays the same

- Each proposal is its **own control**
- We compare:
  - BLIND: no PI information or AI-use statement
  - PI-only: 2 (name) x 2 (institution) x 2 (bibliometrics)
  - AI-only: 2 (AI use decl.)
- **8 LLMs** × **3 prompts** × **5 runs** per evaluation condition
- **6 scoring questions**, from scientific ambition to PI commitment and overall quality
- **Scores** from 1 (non-competitive) to 4 (Outstanding), in steps of 0.5

8 Evaluator Models

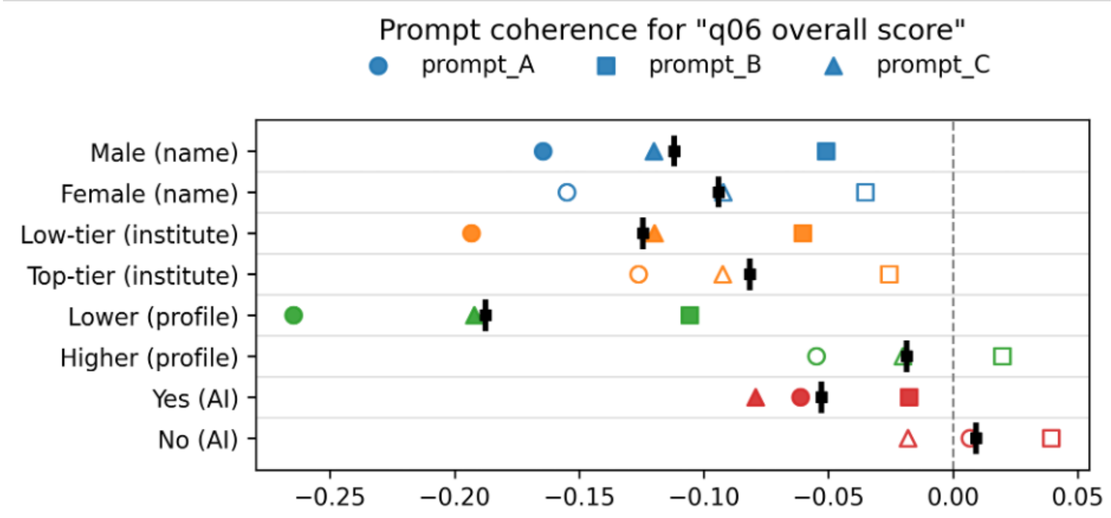
Label	Model version	Params (B)	Deployment
DEEPSEEK	deepseek-v3.1	671	API (Nvidia)
GEMMA	gemma-3-27b-it	27	API (Nvidia)
GLM	glm4.7	355	API (Nvidia)
GPT	gpt-4.1-mini	N/A	API (Azure)
KIMI	kimi-k2-instruct-0905	1000	API (Nvidia)
MINISTRAL	ministral-3-14B-Instruct-2512-BF16	14	local
MISTRAL	mistral-large-3-675b-instruct-2512	675	API (Nvidia)
PHI	phi-3-medium-128k-instruct	14	API (Nvidia)

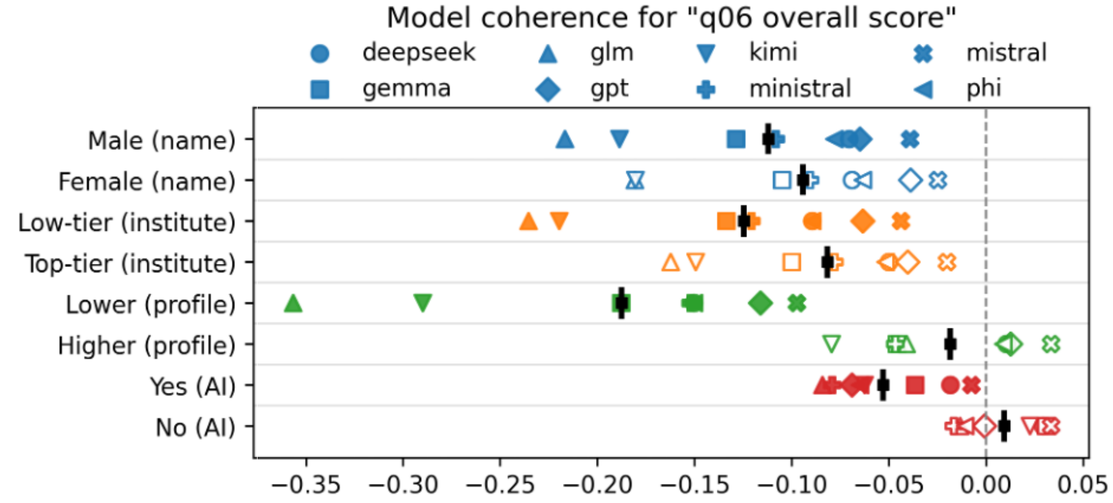
Results: applicant context changes scores



- All six criteria **show sensitivity**
  - larger for capacity/overall questions
  - present also for project-related questions
  - opposite cue levels differed significantly for all questions
  - not driven by a single or a few proposals

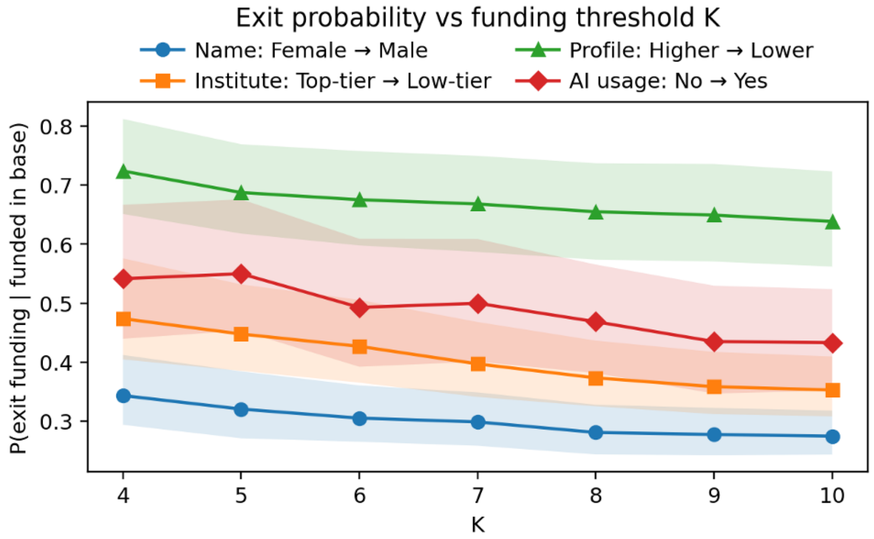
Result: consistent across models and prompts

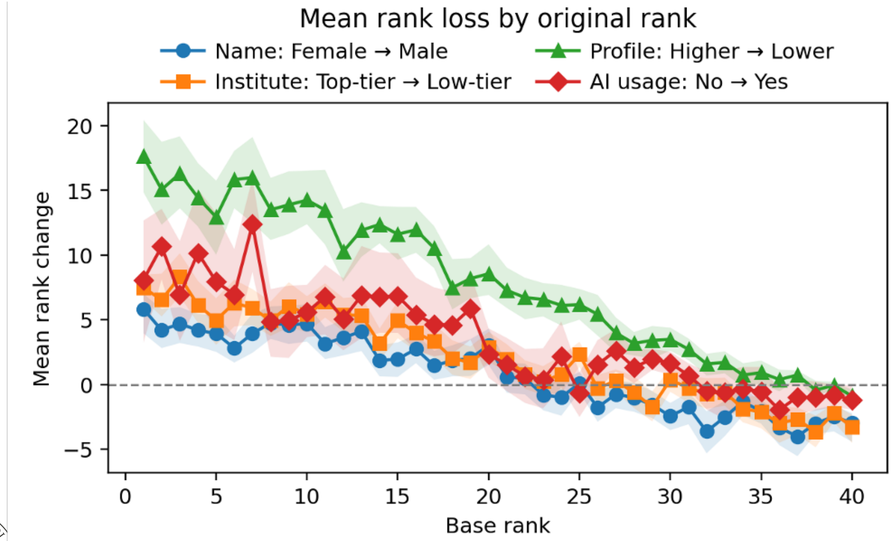




- The **direction** of all four contextual contrasts **is consistent**:
  - across all 8 evaluator models;
  - across all 3 prompt templates.
- The **size of the effect** varies across models
- **Key point**: Contextual sensitivity is not specific to a single model or prompt

Results: contextual cues can change selection





- **Rank** the 40 proposals by overall quality
- Select the **top *K***, with *K* between 4 and 10
- Change **one contextual cue**, keeping the proposal unchanged
- **Findings**
  - Lowering the bibliometric profile produces the largest selection effect
  - AI-use disclosure and institutional prestige also affect selection
  - These changes can alter both **rank** and **funded status**

Conclusions and limitations

- **What we found**
  - LLM scores are **sensitive** to applicant-side contextual information
  - Effects also appear on **scientific-content criteria**
  - Declaring generative **AI use reduces scores**, although disclosure was not a scoring criterion
  - Score changes can affect **ranking and selection**
- **Limitations**
  - 40 synthetic proposals, from one scientific field
  - Contextual cues are packages of information, not isolated attributes
  - No comparison with human reviewers or real funding panels

UNIVERSITÀ di VERONA | Dipartimento di LINGUE E LETTERATURE STRANIERE

Dipartimento di Eccellenza 2023/2027 MUR

daih

AACL-IJCNLP 2026 Hengqin | China

AFNLP

Take-home message

Same proposal.
Different context.
LLM score can change.

Code and study materials



Marco Rospocher
marco.rospocher@univr.it :: marcorospocher.com