Trust in AI in research: Yet another variable

Olya Vvedenskaya

10 April 2026

Olya Vvedenskaya, MD, PhD

Academia

- MD RNRMU, Moscow, Russia &
University of Pittsburgh, USA - 2012

- PhD Humboldt University, MDC, Charite, Berlin, Germany - 2017

- Postdoc MPI-CBG, Dresden, Germany - 2017-2021

Industry

- Manager in Spectroswiss, Switzerland - 2021-2022

- Communications Manger, Lipotype, Dresden, Germany - 2022 – now

Non-profit sector

- Co-founder, Dragonfly Mental Health, USA - 2019-2025

- Co-founder, community lead, outreach lead, Sci.STEPS, Berlin, Germany - 2022 - now

Definitions

Trust is commonly defined as a willingness to be vulnerable based on expectations of competence, integrity, and benevolence (Svare, 2020 Baghramian, 2020).

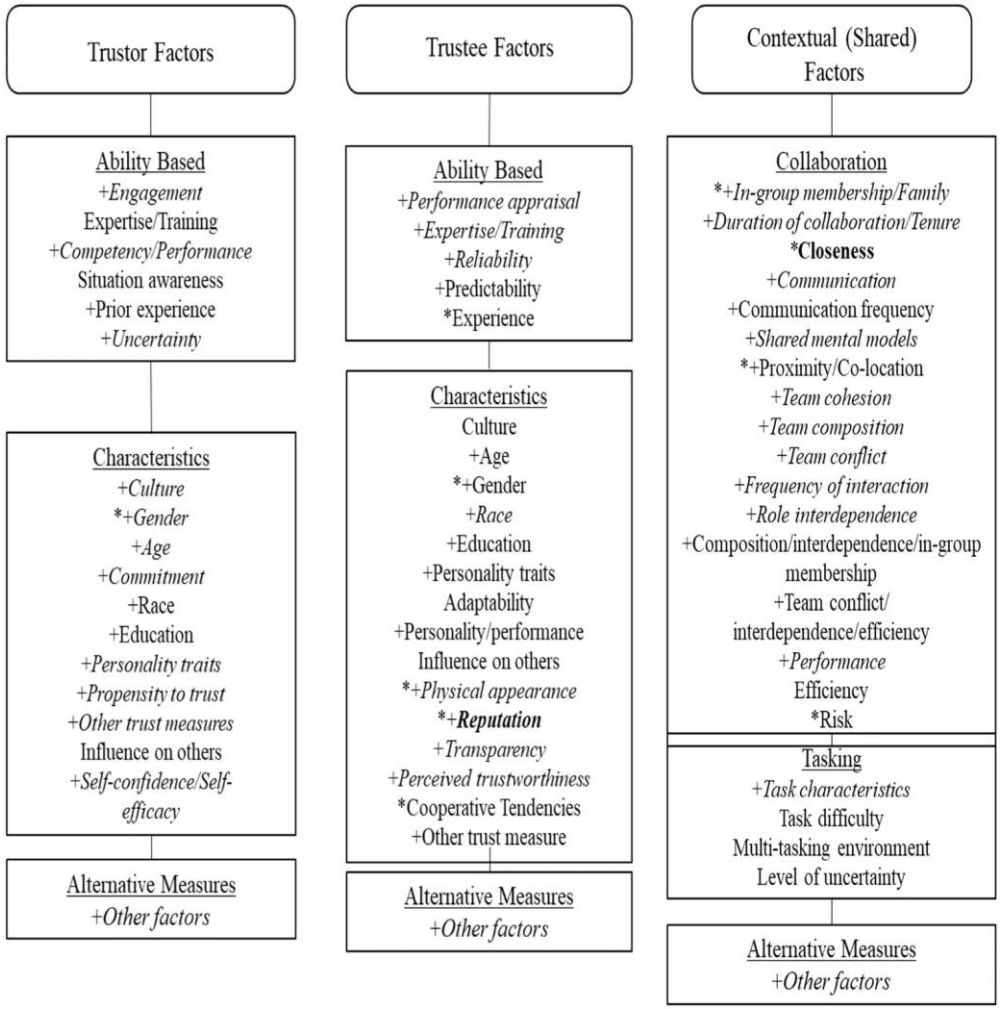
Since this definition emerged from research on **human-to-human** relationships, applying it to AI requires careful consideration (McGrath, 2024).

Human-AI trust depends less on social cues or shared relationships and more on system performance, transparency, ethical design, and the context in which the system is used.

Olya Vvedenskaya
warenje.net

Understanding human-to-human trust

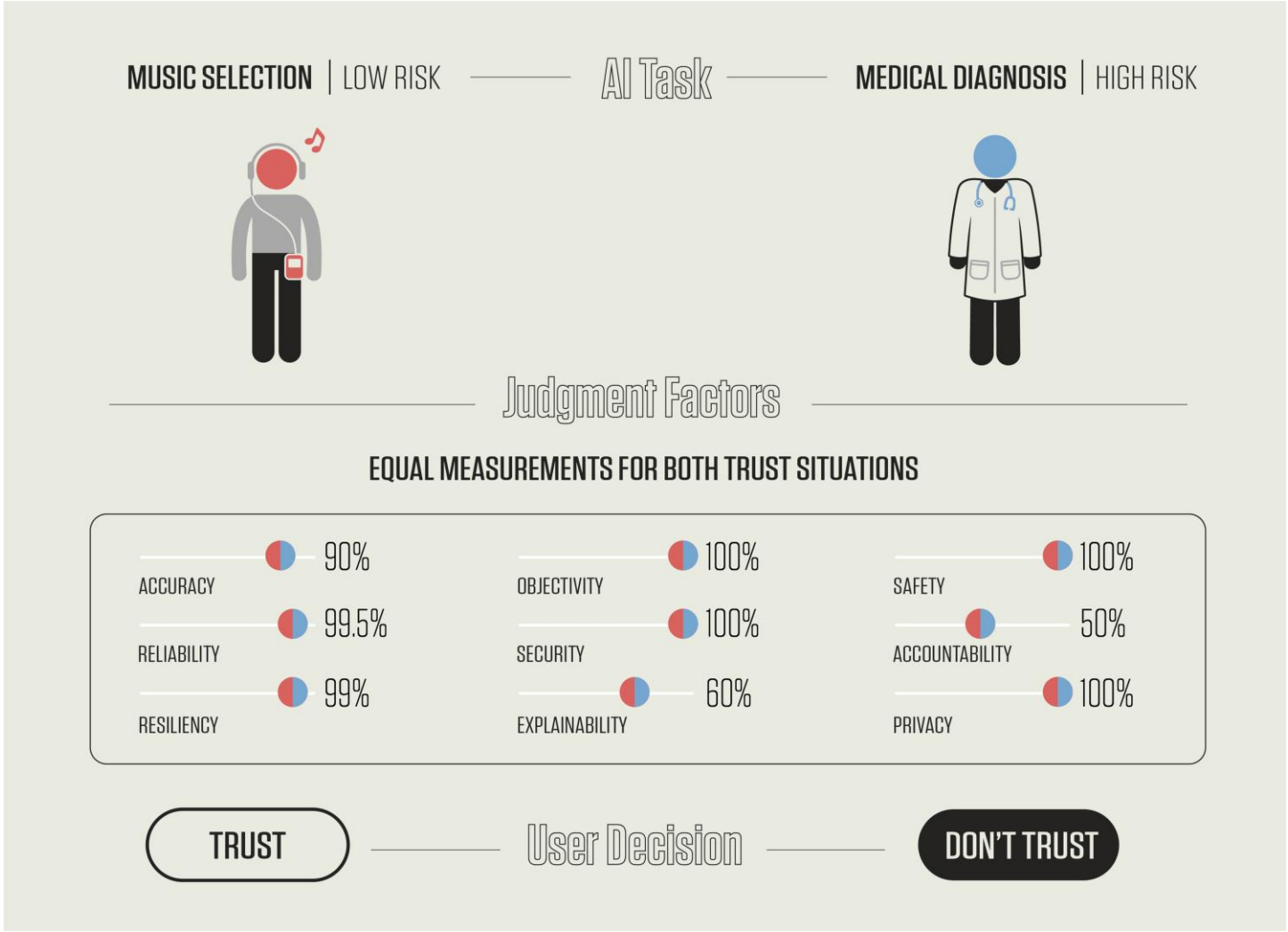
Trustor (the person that trusts) and the trustee (the one who is to be trusted)



Model is
revised by Hancock at al. (2023)
from Mayer et al. (1995).

Olya Vvedenskaya
warenje.net

Understanding human-AI trust



Brian Stanton, NIST, 2021

Olya Vvedenskaya
warenje.net

Measuring and/or understanding?

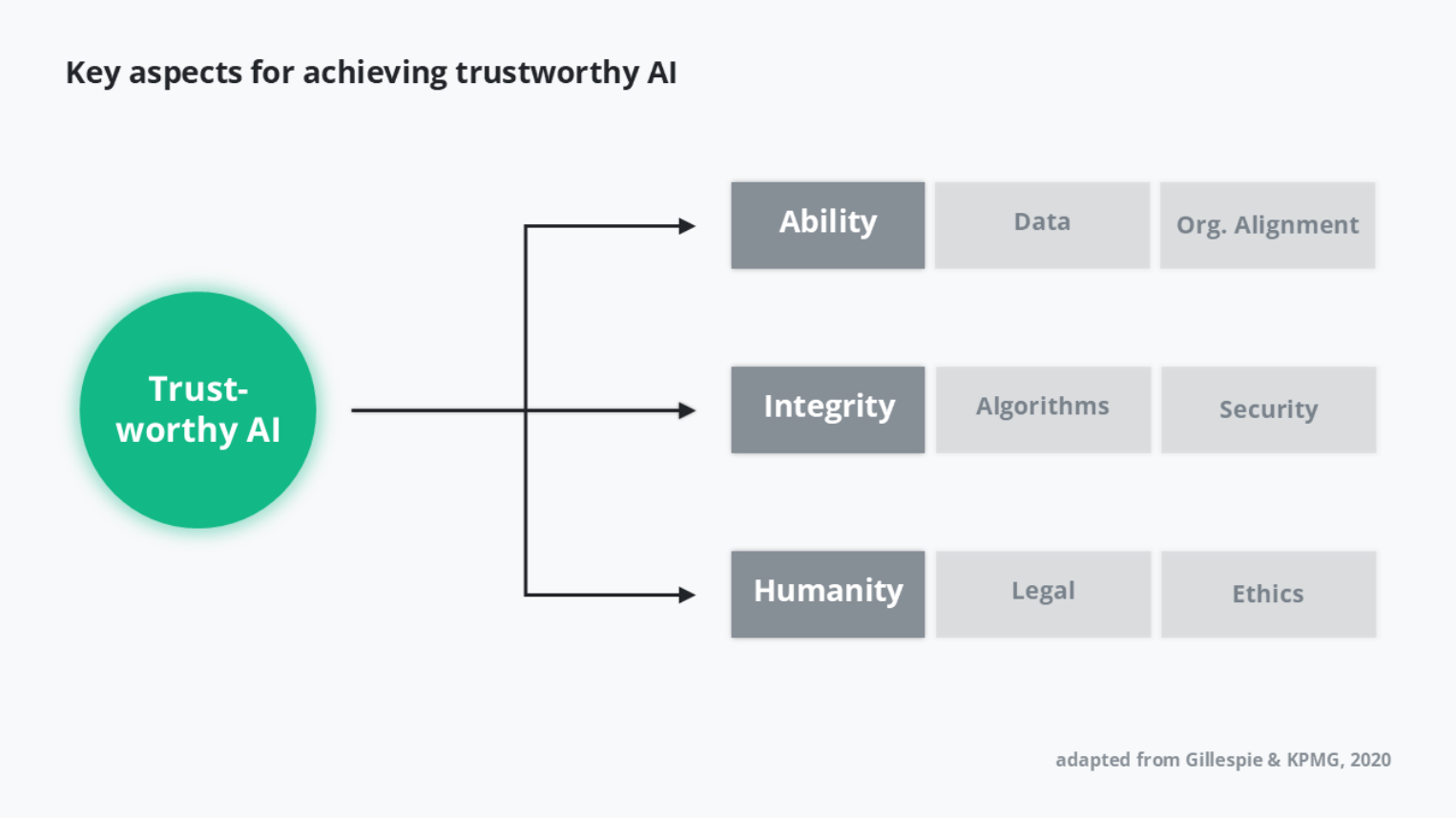
- What metric(s) can we use in human-AI trust? Are there any reasonable metrics at all?

- Are they field- and application-specific?

- How can we use this knowledge in helping researchers navigate the jungle of existing tools?

Olya Vvedenskaya
warenje.net

Core principles of trustworthy AI



Additional readings:
Blau et al. 2024
ALLEA Code of Conduct
Chen et al. 2024

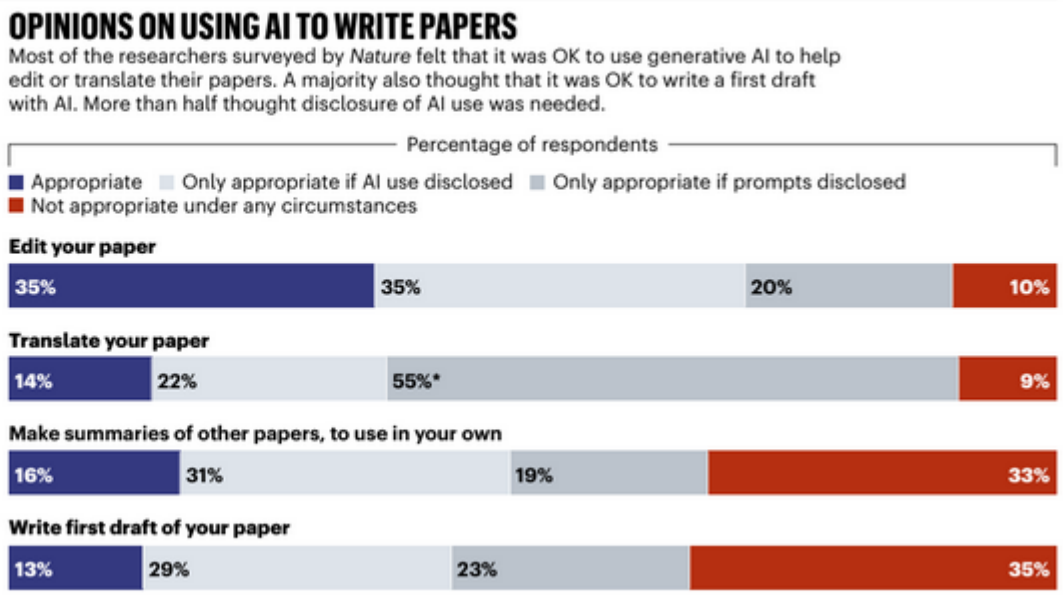
Olya Vvedenskaya
warenje.net

Feature

SCIENTISTS SPLIT ON ETHICS OF AI USE

A *Nature* survey of 5,000 researchers finds contrasting views on when it's acceptable to involve AI in research papers and what needs to be disclosed.

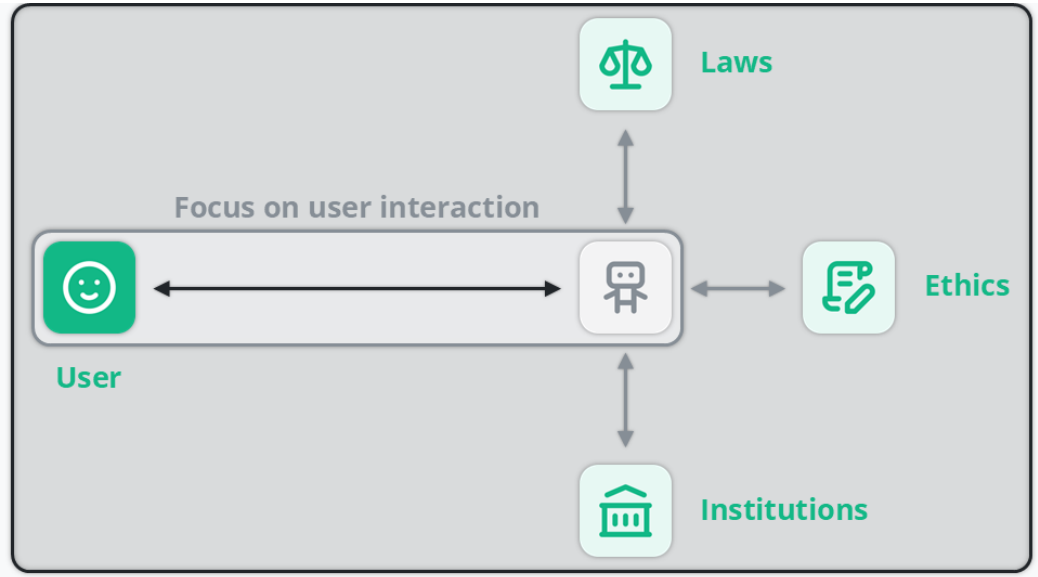
By Diana Kwon



Olya Vvedenskaya
warenje.net

Further aspects

- Engaging multiple stakeholders and levels

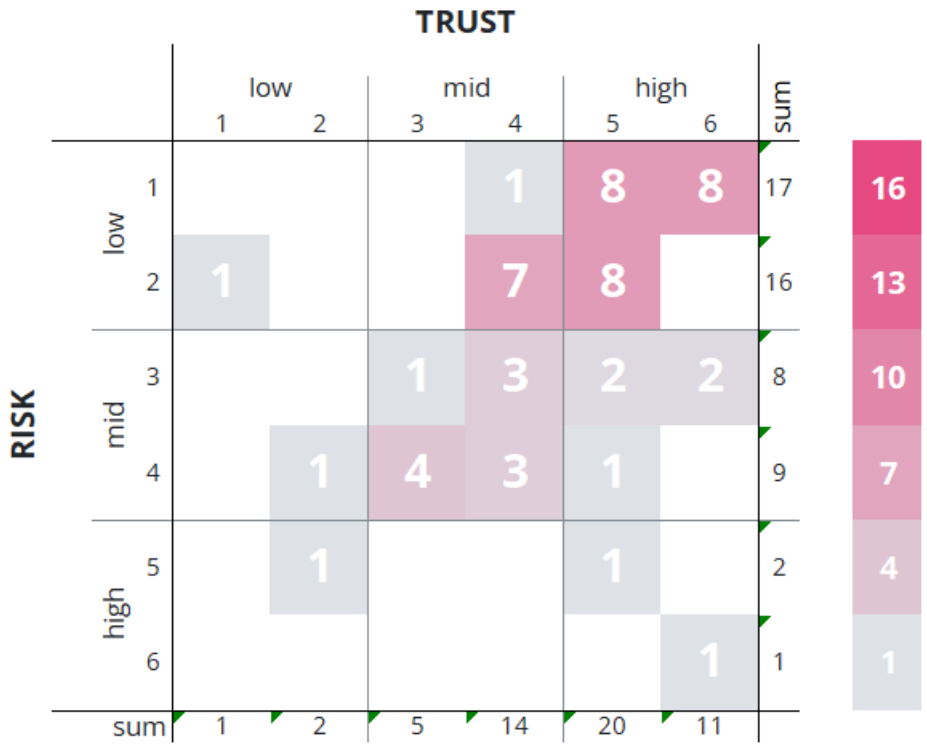


- Dehumanizing AI
- Balancing trust appropriately and taking a pulse - surveying researchers' trust in AI

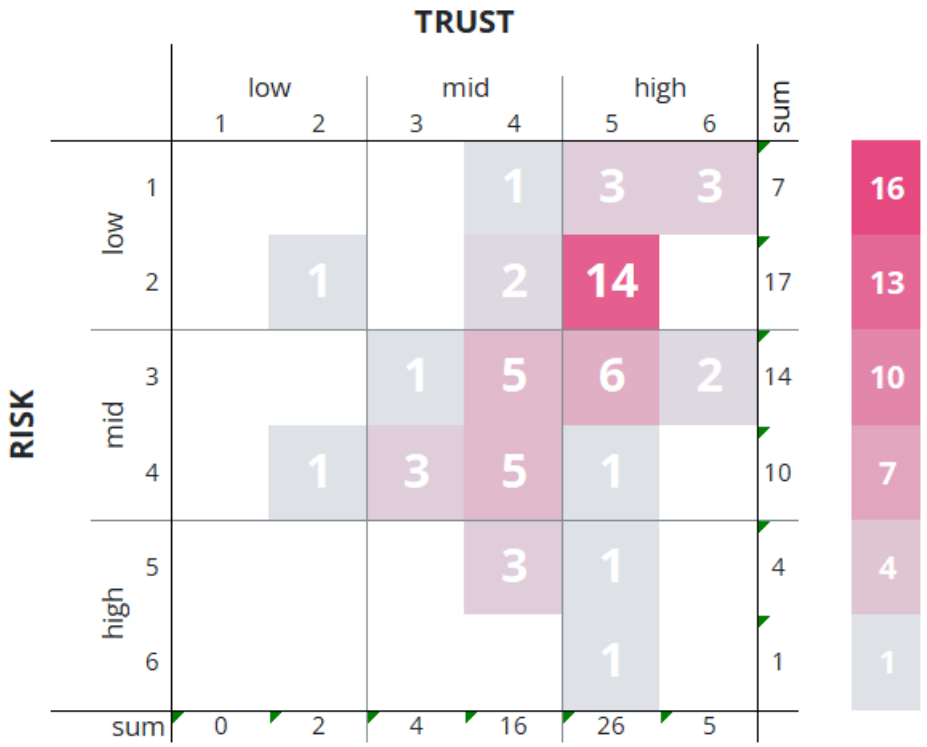
Olya Vvedenskaya
warenje.net

Low-risk and high-trust activities

Searching scientific literature

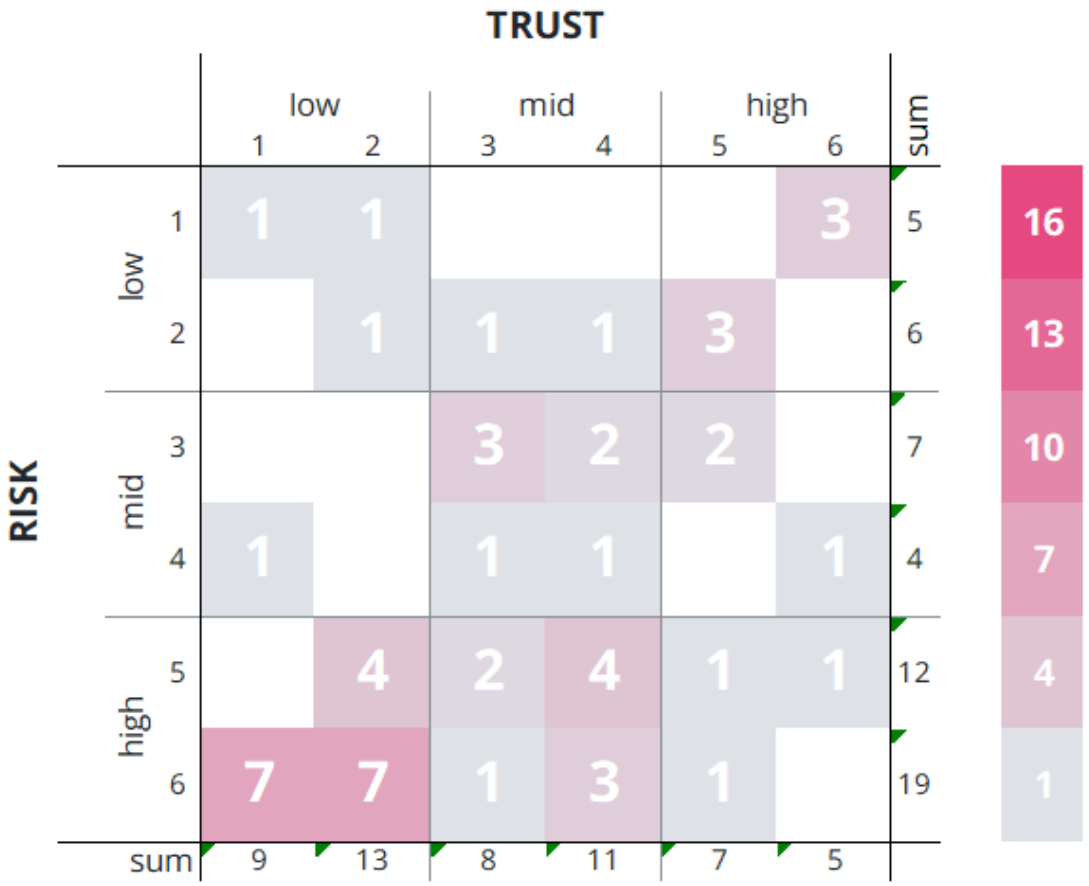


Summarizing papers



Olya Vvedenskaya
warenje.net

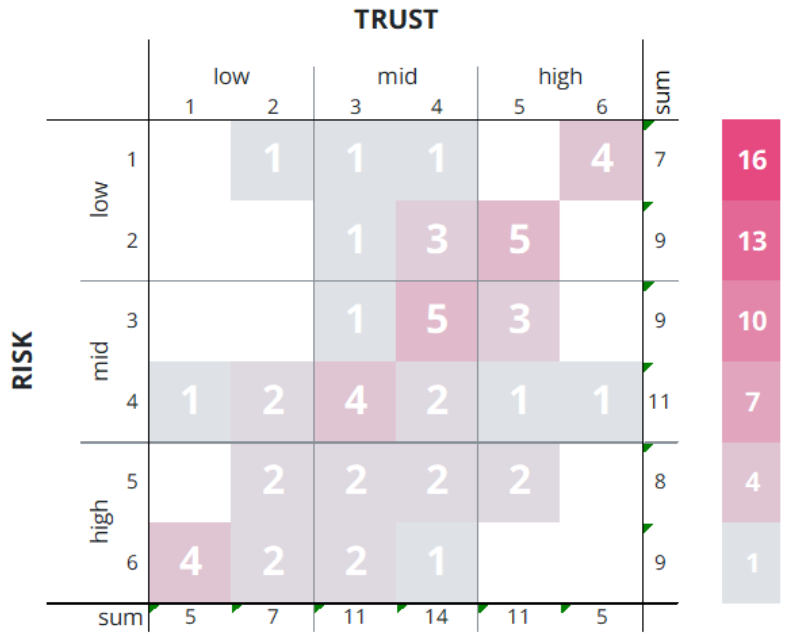
High-risk and low-trust activities



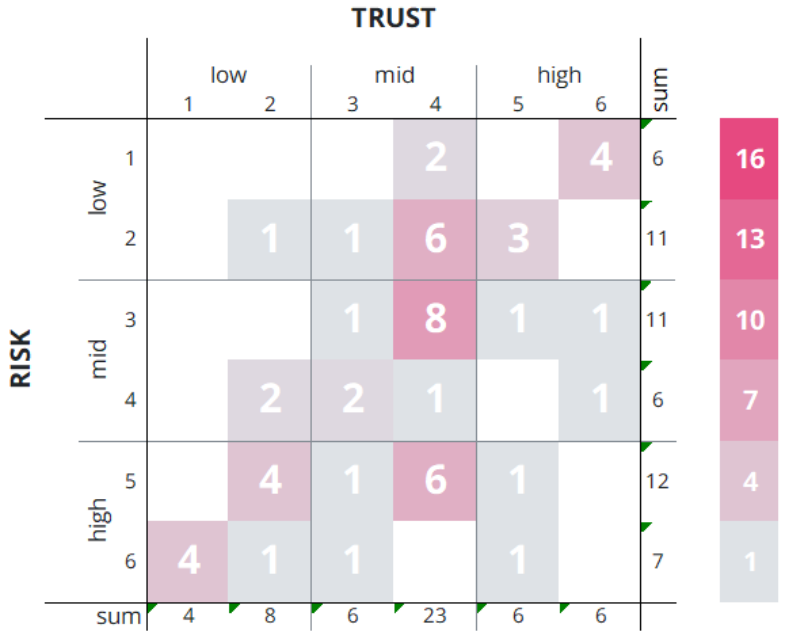
Olya Vvedenskaya
warenje.net

Highly subjective tasks

Suggesting hypotheses

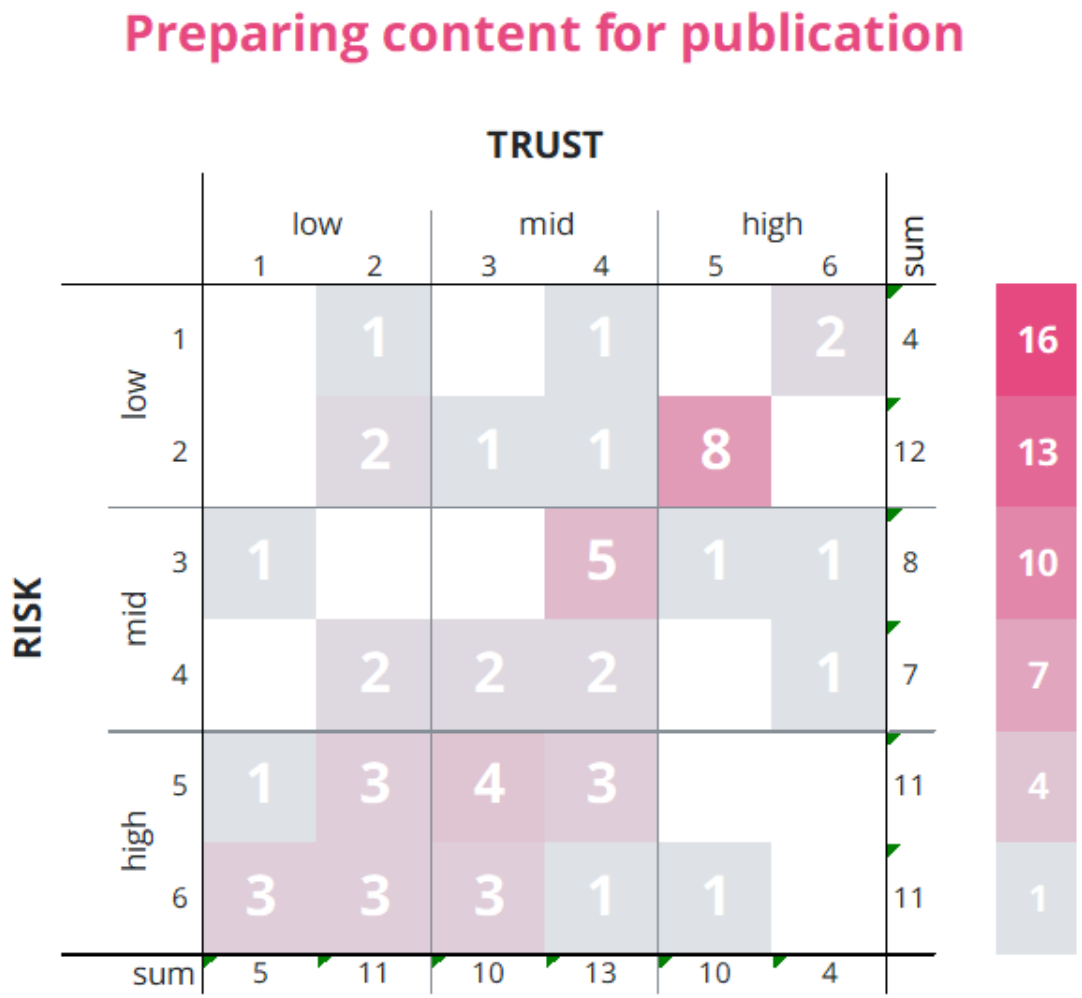


Drafting or editing research reports



Olya Vvedenskaya
warenje.net

Highly subjective tasks



Olya Vvedenskaya
warenje.net

Conclusions

- Trust in AI changes depending on the task and the level of perceived risk
- Scientists may confidently use scienceOS for some tasks while remaining cautious about others
- Transparency, clear system limits, and human oversight help create well-founded trust
- Appropriate trust, rather than blind trust, is essential for responsible AI use in scientific research
- When using any AI tool: do the background check, understand why and how you are doing it, evaluate the output, disclaim the AI usage

Olya Vvedenskaya
warenje.net

Thank you!

… so many people to thank!

Questions?

Staying in touch:
olga.m.vvedenskaya@gmail.com